THE UNIVERSITY
of EDINBURGH

A genetic and epigenetic editing approach to characterise the nature and function of bivalent histone modifications

Ailbhe Brazel

Doctor of Philosophy - The University of Edinburgh - 2017

This thesis is dedicated to my brother Sam.

Thesis Declaration

Name of Candidate:	Ailbhe Brazel	*UUN*	S1360165
University email:	ailbhe.brazel@roslin.ed.ac.uk		

Degree Sought:	Doctorate of Philosophy	*Number of words in the main text of Thesis:*	59,294
Title of Thesis:	A genetic and epigenetic editing approach to characterise the nature and function of bivalent histone modifications		

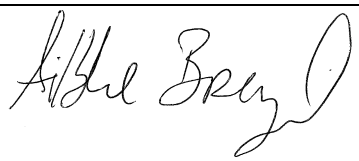
I certify:

(a) that the thesis has been composed by me, and

(b) either that the work is my own, or, where I have been a member of a research group, that I have made a substantial contribution to the work, such contribution being clearly indicated, and

(c) that the work has not been submitted for any other degree or professional qualification except as specified, and

(d) that any included publications are my own work, except where indicated throughout the thesis and summarised and clearly identified on the declarations page of the thesis.

Signature:	

Abstract

In eukaryotes, DNA is wrapped around a group of proteins termed histones that are required to precisely control gene expression during development. The amino acids of both the globular domains and unstructured tails of these histones can be modified by chemical moieties, such as methylation, acetylation and ubiquitination. The 'histone code' hypothesis proposes that specific combinations of these and other histone modifications contain transcriptional information, which guides the cell machinery to activate or repress gene expression in individual cell types.

Chromatin immunoprecipitation (ChIP) experiments using undifferentiated stem cell populations have identified the genomic co-localisation of histone modifications reported to have opposing effects on transcription, which is known as bivalency. The human α-globin promoter, a well-established model for the study of transcriptional regulation, is bivalent in embryonic stem (ES) cells and this bivalency is resolved once the ES cells terminally differentiate (i.e. only activating or repressing marks remain). In a humanised mouse model, the deletion of a *bone fide* enhancer within the human α-globin locus results in heterogeneous expression patterns in primary erythroid cells. Notably, this correlates with an unresolved bivalent state at this promoter in terminally differentiated cells.

Using this mouse model it is not feasible to ascertain whether the transcriptional heterogeneity observed in the cells lacking an α-globin enhancer is reflective of epigenetic heterogeneity (i.e. a mixed population of cells) rather than co-localisation of bivalent histone modifications within the same cells. Furthermore, the functional contribution of bivalency to development has yet to be described. To address these difficulties, I aimed to generate a fluorescent reporter system for human α-globin to facilitate the separation of transcriptionally heterogeneous erythroid cells. This model will provide material for ChIP studies on transcriptionally active and inactive populations to determine whether the epigenetic bivalency is reflective of a mixed cell population or true bivalency. In addition, I aimed to produce epigenetic editing tools to target bivalent promoters, which in combination

with *in vitro* differentiation assays would provide an interesting framework to test the function of bivalency during development.

In this study, I extensively tested gene-editing strategies for generating a fluorescent reporter knock-in in humanised mouse ES cells. I validated the suitability of humanised mouse ES cell lines for gene targeting studies and optimised a robust *in vitro* differentiation protocol for studying erythropoiesis. I utilised both recombineering and CRISPR/Cas9 gene editing tools in tandem with *PiggyBac* transposon technology, to knock-in the reporter gene. I made significant steps in gene targeting and successfully inserted the reporter downstream of the α-globin gene. I also generated a cloning system to express site-specific DNA-binding domains (TALEs) fused to epigenetic regulators with the aim to resolve bivalent histone modifications *in vitro*. From preliminary tests using these fusion proteins targeting *Nrp1,* a bivalent promoter in mES cells, I observed mild but significant changes in gene expression although histone modifications were unchanged. The various tools generated and tested in this study provide a solid foundation for future development of genetic and epigenetic editing at the human α-globin and other bivalent loci.

Lay Abstract

Mammals are made up of a large number of very different cell types that come together to form a complete organism. It is important to note that every cell in an individual organism has an identical copy of DNA. This DNA acts as a blueprint and contains information, encoded in units called genes, on how to generate every protein required in an organism. The proteins act as building blocks and particular sets of proteins are required to construct each different cell type. To ensure the correct set of proteins is available, each cell must turn 'on' (activate) the correct set of corresponding genes while also ensuring the rest of the genes remain switched 'off' (repressed).

The DNA is wrapped around a group of proteins, termed histones, which are involved in deciding what genes get activated or repressed in an individual cell. The activating or repressing action of histones involves different chemical modifications, often called epigenetic marks. A number of genes have been identified that appear to be marked by modifications associated with both gene activation and repression, a so-called 'bivalent' state. However, it is possible that in some cases of bivalency only the cells that do express a gene contain the modification associated with gene activation and *vice versa*.

In testing this hypothesis I aimed to generate a new humanised mouse model which could facilitate the visualisation and separation of individual cells based on the expression of an important bivalent gene, human α-globin. Mutations that dysregulate α-globin cause α-thalassemia, a major blood disorder. In this study I successfully completed the initial steps in generating this model.

There is not enough evidence to show if histone modifications instruct the cell to make decisions about gene activation or repression or if the modifications act more like an indexing system, preserving the information on whether a gene is activated or repressed. To test the precise function of these modifications I generated a system to construct a number of tools that could accumulate or remove individual histone modifications at particular genes. Interestingly, one of the tools was able to induce a mild activation of the target gene. The tools generated and tested in this

study provide a solid foundation for investigating the nature and function of histone modifications at human α-globin and other genes. The analysis of how gene expression is regulated will aid in the greater understanding of how both normal and diseased cell types develop.

Acknowledgements

Firstly, I would like to thank my supervisors. To Doug Vernimmen, I am grateful for giving me the opportunity to work on this project and coming to this wonderful city. To Peter Hohenstein, I am grateful for making me an unofficial lab member and telling me there's no such thing as a stupid question. To Paul Digard, I am grateful for rescuing me from doom and disaster on the (many) occasions I needed you. I'm sure nobody will miss those long Monday morning meetings but I really want you all to know how much I appreciate the time you put into this project. Thanks also to the other members of my thesis committee, Helen Sang and especially Andy Law, who provided much advice and words of support when needed.

Next, well it's a long list! They say it takes a village to raise a PhD student and I've certainly had many wonderful lab mammies and daddies taking care of me these past few years. In the beginning on the 2nd floor many people helped with my daily what-am-I-doing moments, particularly Bill Gregory. Iveta Gazova, it was always a pleasure (who am I kidding, we were both in pain!) to learn the ropes with you. After moving down to the 1st floor I was delighted to meet a whole new lab family. Charis Hogg, your patience and kindness was such a comfort and help when I was feeling low. I could not have survived without the completely random, and sometimes incredibly helpful, lab chat from Barry Bradford, Claire Stenhouse, Dec King, Sam Eaton, Kirsty Ireland, Andrew Castle and many many more. Downstairs I have to thank poor Graeme Robertson, Bob Fleming and Anna Raper for my many Flow and imaging questions, and all of the guys in stores and the CSU. You guys are so incredibly patient and helpful; the institute is lucky to have you all! Thank you also to the students I briefly supervised who helped me with some of the experiments in this thesis; Sarah Mounedji, Eirini Kallimasioti and Justin Auerbach.

Then comes the most emotional lab room of them all; the ES tissue culture room! Here I have to give a huge thank you to Derya Ozdemir who helped me with all things CRISPR and tissue culture and was also an excellent desk buddy for the last four years; I will miss you! Another huge thank you must be given to Linda Sutherland for her patience and constant entertaining stories. Then a big laugh and hug for all the other members of the ES tissue culture room therapy group; Rocio

Rojo, Anagha Krishna, Shahida Sheraz, Chris Cartlidge, Stephen Meek, Tom Burdon. Suffering and complaining together in that tiny room helped us all survive.

I think it's now time to thank all the wonderful friends I made in my PhD time that helped me to survive. The solid "lunch group" from our first year remained a support group we could always rely on. From this group, I'm so grateful to have made some (Ph)rien(D)s for life! Anna Mantsoki, I will always love your resting angry face and you were so sorely missed this last year late nights in Roslin; looking forward to making up lost time on holidays soon. Emma Hurst, I am so delighted we met on Day 1 and now I can't wait to decorate my room in your new flat. Heather Mathie, even though they all left us behind as the last ones to finish I'm so glad we were there together as I couldn't have done it without our moans. Abi Bremner and Craig Johnson, your cynicism and randomness made me laugh constantly, thank you. Thanks also to my flatmates over the years for the times we had relaxing and moaning together, particularly Selene Jarrett. I'll be seeing you all again very soon.

My wonderful family also need a huge thank you. My parents have given me endless support throughout this entire process. Their work ethic and curiosity was part of what inspired me to do a PhD in the first place and their kindness helped me to survive it. Thank you both. My brothers and sister, Matthew, Sam and Emma, all provided laughter, kind words and advice when needed, thank you.

Finally, I get to thank Diarmuid. You inspire me to look for interesting questions, design simple tests and thoughtfully analyse the results. I can never thank you enough for all the emotional support, the innumerable midnight-3am flights, the constant discussions about all our experiments and helping me get through these last four years and this thesis. You are everything. Can you believe we did it? I can't wait to move into the next, hopefully more relaxing (!), chapter of our lives together.

List of Frequently Used Abbreviations

Amp	Ampicilin
BAC	Bacterial artificial chromosome
Cas9	(Clustered regularly interspaced short palindromic repeat)-associated protein 9
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation with next generation sequencing
CIP	Calf intestinal alkaline phosphatase
CMV	Cytomegalovirus
co-ChIP-seq	Combinatorial indexed chromatin immunoprecipitation with next generation sequencing
CpG	5`-Cytosine-phosphate-guanine-3`
CRISPR	Clustered regularly interspaced short palindromic repeat
Ct	Cycle threshold
dCas9	Catalyticaly dead Cas9
ddH_2O	Double distilled H2O
Dex	Dexamethasone
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Dnmt	DNA methyltransferase
dNTP	Deoxyribonucleotide triphosphate
DSB	Double stranded break
dsDNA	Double stranded DNA
E. coli	*Escherichia coli*
EB	Embryoid body
EB Diff	EB differentiation
EDTA	Ethylenediaminetetraacetic acid
Eed	Embryonic ectoderm development
ES	Embryonic stem

Ex	Exon
Ezh2	Enhancer of zeste 2
FACS	Fluorescent activated cell sorting
FCS	Foetal calf serum
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
gDNA	Genomic DNA
Gent	Gentamycin
GFP	Green fluorescent protein
GOI	Gene of interest
H2Aub	Mono-ubiquitination of lysine 119 on histone 2A
HBA	Human alpha globin
HDR	Homology directed repair
Hn	Histone n
HS-40	DNaseI hypersensitive site-40
Indel	Small insertion or deletion
JMJD	Jumonji-domain containing proteins
Kan	Kanamycin
Kan^R	Kanamycin resistant
kb	kilobase
mCherry	Monomeric cherry
mES cell	Mouse embryonic stem cell
mRNA	Messenger ribonucleic acid
Mut	Mutation
NFW	Nuclease free water
NHEJ	Nonhomologous end joining
PBS	Phosphate buffered saline
PcGs	Polycomb group proteins
PCR	Polymerase chain reaction
PGK	Phosphoglycerate kinase
pTFE	TALE fusion expression vector
qPCR	Quantitative real-time polymerase chain reaction
Rbbp5	Retinoblastoma-binding protein 5

RMCE	Recombinase mediated cassette exchange
RNA	Ribonucleic acid
rSAP	Recombinant shrimp alkaline phosphatase
RTS	Recombination target site
RVD	Repeat-variable diresidue
SNP	Single nucleotide polymorphism
TF	Transcription factor
tNrp1	TALE array targeting the promoter of *Nrp1*
UTR	Untranslated region
WTH	Wild-type humanised

Table of Contents

Table of Figures

Table of Tables

Chapter 1 Introduction

1.1 Epigenetics and transcriptional regulation

Epigenetics traits are heritable phenotypes that are not derived from changes in DNA sequence (Berger et al., 2009). Prof. Conrad Waddington, a biologist and a philosopher who held the Directorship of the Institute of Genetics in the University of Edinburgh, coined this term in the 1940s (Slack, 2002). Waddington was the first to describe an "epigenetic landscape" in which each embryonic stem (ES) cell makes numerous decisions that can lead it down different developmental pathways.

We now have a much clearer insight into the molecular causes of these epigenetic traits and developmental decisions. In a cell, DNA is wrapped around packaging proteins called nucleosomes in a structure termed "chromatin". Nucleosomes are composed of four proteins called histones which form an octamer through the interaction of a tetramer of two copies of H3 and H4 with two dimers of H2A and H2B. Both the DNA and the histones can be chemically modified by methylation, while the histones can also be acetylated, phosphorylated and ubiquitinated. A set of enzymatic complexes are involved in the deposition of these chemical groups. These complexes include a "reader" domain that facilitates the correct localisation of the mark and a "writer" domain which catalyses the modification itself. Many "eraser" enzymes have also been identified that catalyse the removal of these modifications. Combinations of these modifications form a complex epigenetic code that is interpreted by the cell machinery. In synergy with both the general and cell-specific transcription factors (TFs), this code can influence the transcriptional activity of underlying genes. Many of these modifications are stably inherited by daughter cells after cell division. Accordingly, these DNA and histone modifications are often called epigenetic modifications. In this section, I will discuss the epigenetic mechanisms involved in the activation and repression of gene transcription, with a particular focus on histone methylation.

1.1.1 Epigenetics and gene activation

Proteins associated with DNA (such as chemically modified histones), can be isolated by chromatin immunoprecipitation (ChIP) experiments. The genomic localisation of proteins can be interrogated by ChIP combined with qPCR (ChIP-qPCR) or next generation sequencing (ChIP-seq). A number of chromatin modifications, such as the acetylation of a vast number of lysine residues on histone tails, have been correlated with gene activation (Wang et al., 2008). Histone acetylation reduces the positive charge of histones, therefore weakening the interactions between the histones and the negatively charged DNA (Bannister and Kouzarides, 2011). This relaxed structure is called "euchromatin" and allows transcriptional machinery to access the DNA more easily (Wang et al., 2008). Histone phosphorylation at a number of residues has also been linked to gene activation, as well as chromatin compaction during mitosis and meiosis (Rossetto et al., 2012). The methylation of certain histone amino acids has likewise been correlated with gene activation, for example, tri-methylation of lysine 4 on histone 3 (H3K4me3) and H3K36me3. H3K36me3 is catalysed by the histone methyltransferase SET2, and is highly enriched across the entire gene body in actively transcribed genes (Bannister et al., 2005; Krogan et al., 2003; Xiao et al., 2003). Below I will focus on H3K4me3, and discuss the evidence that suggests this mark may have a role in gene activation.

1.1.1.1 H3K4me2/3 and gene activation

One of the most intensively studied histone modifications is H3K4me3. H3K4me3 is enriched close to the transcriptional start site (TSS) of active promoters, which has long suggested it is involved in transcriptional activation (Bernstein et al., 2005; Santos-Rosa et al., 2002). The breadth of H3K4me3 domains is positively correlated with consistency in gene transcription at a single cell level (Benayoun et al., 2014). The targeted writing of H3K4me3 at specific promoters using epigenetic editing tools (described in detail in section 1.5.2, pp. 40) has shown that an increase of localised H3K4me3 is sufficient to cause a mild increase in gene activation at tested gene promoters (Cano-Rodriguez et al., 2016).

SET domain containing proteins, including SET1 in yeast or SET1-like proteins in mammals (SET1A, SET1B, MLL1, MLL2, MLL3 and MLL4), catalyse the deposition of H3K4me2/3 (Herz et al., 2013; Shilatifard, 2012). H3K4me2 has also been associated with active gene transcription (Meissner et al., 2008). Most of these histone methyltransferase complexes contain zinc finger CxxC DNA binding domains which bind to unmethylated regions of DNA that are enriched for the CpG dinucleotide, termed CpG islands (Blackledge et al., 2013; Thomson et al., 2010). Vertebrate CpG islands are ~1 kb in length and are high in GC content. They are found in nearly all housekeeping genes and certain developmental regulators and tissue-specific genes (Deaton and Bird, 2011). Cfp1 is the CxxC containing component of the Set1d complex (Thomson et al., 2010). The binding specificity of Cfp1 to unmethylated CpG islands helps prevent 'leakage' of H3K4me3 to ectopic sites, but has not been proven necessary for histone methylation at target genes or critical for target gene expression regulation (Clouaire et al., 2012). Importantly, the deposition of H3K4me3 at artificial, promoterless CpG islands *via* Cfp1 is not sufficient to recruit the RNA polymerase II complex (PolIII), suggesting that the recruitment of these two complexes occurs independently (Thomson et al., 2010). Cfp1 also contains PHD domains which recognise and bind to H3K4me1/2/3, promoting further deposition of the mark (Eberl et al., 2013). The MLL complexes appear to be important in methylating H3K4 at specific subsets of target sites, such as homeotic (Hox) genes and enhancers (Howe et al., 2017).

A number of different studies have attempted to show a functional link between H3K4me3 and gene activation. Primarily, the TAF3 subunit of Transcription Factor II D (TFIID) binds to H3K4me3 through PHD finger interactions (Lauberth et al., 2013; Vermeulen et al., 2007). The recruitment of the TFIID in turn facilitates the assembly of the PolIII preinitiation complex (Lauberth et al., 2013; Vermeulen et al., 2007). H3K4me3 also binds a number of other epigenetic effector proteins including the NURF chromatin remodelling complex and histone acetyl transferases, which act to facilitate chromatin accessibility and gene transcription (reviewed in Wozniak & Strahl 2014). Finally, H3K4me3 has also been implicated in the prevention of gene repression. JMJD2A, a H3K9me3 demethylase, binds to H3K4me3 through its double Tudor domains to remove this histone mark

that is associated with gene repression (Huang et al., 2006; Klose et al., 2006). The presence of the H3K4me3 also prevents the binding of the *de novo* DNA methylase Dnmt3a, thus preventing permanent repression *via* DNA methylation (Zhang et al., 2010).

It is clear from the evidence discussed above that H3K4me3 is positively correlated with gene expression (Benayoun et al., 2014; Bernstein et al., 2005; Cano-Rodriguez et al., 2016; Santos-Rosa et al., 2002) and certain studies have even identified mechanistic links between H3K4me3 and the activation of gene transcription (Huang et al., 2006; Klose et al., 2006; Lauberth et al., 2013; Vermeulen et al., 2007; Wozniak and Strahl, 2014; Zhang et al., 2010). Despite this body of evidence, a number of studies have shown that the removal of the H3K4me3 methyltransferase complex members, and a global reduction in H3K4me3, does *not* result in major changes in global gene expression levels (Clouaire et al., 2012; Hormanseder et al., 2017; Lenstra et al., 2011; Muramoto et al., 2010). Furthermore, genes marked with high levels of H3K4me3 (and no detectable marks associated with gene repression) are not always expressed (Vastenhouw et al., 2010). Instead of directly activating expression, H3K4me3 may instead be important in preserving the memory of transcriptional activation during cell division (Hormanseder et al., 2017; Howe et al., 2017; Muramoto et al., 2010). In agreement with this hypothesis, live cell imaging in *Dictyostelium* cells using the MS2 system for visualising transcription in real time (described in section 1.5.2.1, pp. 40) was performed on dividing cells (Muramoto et al., 2010). This study revealed that knocking out components of the Set1 complex in these cells led to a loss in H3K4me2/3 levels and a loss of the inheritance of transcriptional state (Muramoto et al., 2010). Similarly, H3K4me3 is involved in maintaining the transcriptional memory of donor cell identity in nuclear transfer experiments in *Xenopus* embryos (Hormanseder et al., 2017). A reduction of H3K4me3 levels in *Xenopus* embryos during nuclear transfer experiments reduced the levels of donor-cell specific gene transcription and improved the developmental capacity of nuclear transfer embryos (Hormanseder et al., 2017). A reduction of H3K4me3 enrichment can reduce transcriptional heterogeneity in a population and improve consistency in gene expression pathways (Howe et al., 2017).

Indeed, it is clear that H3K4me3 enrichment is associated with gene activity, although it does not appear to be causative in certain cases (Cano-Rodriguez et al., 2016; Lauberth et al., 2013). H3K4me3 may more indirectly influence the maintenance of an active transcriptional state at the majority of genes (Hormanseder et al., 2017; Howe et al., 2017; Muramoto et al., 2010). In the following section I will discuss correlations between epigenetic marks and gene repression.

1.1.2 Epigenetics and gene repression

As previously described, histones that are not acetylated have a strong positive charge resulting in strong interactions with the negatively charged DNA molecules they are wrapped around. This leads to compaction of the chromatin into areas of "heterochromatin" in which the transcriptional machinery is unable to access the DNA, resulting in low expression of the genes located in these regions. Two types of heterochromatin are observed in cells (reviewed in Becker et al. 2016). Constitutive heterochromatin is stably formed in every cell type in an organism and can be found close to telomeres and centromeres. The histone methylation mark H3K9me2/3, deposited by members of the Suv39 family including SETDB1, GLP and G9A, and also PRDM2 is enriched in constitutive heterochromatin (and in patches of facultative heterochromatin). H3K9me2/3 then recruits heterochromatin protein 1 (HP1), which causes chromatin compaction (Becker et al., 2016). Facultative heterochromatin is converted from euchromatin in certain cells, for example the inactivated X chromosome in female mammalian cells. The formation of facultative heterochromatin involves the deposition of several different histone marks which can include H3K27me3, H3K9me2 and the mono-ubiquitination of lysine 119 on histone H2A (H2Aub) (Trojer and Reinberg, 2007).

The methylation of 5-carbon on cytosine residues in CpG dinucleotides of DNA represses transcription (Maeder et al., 2013a; Stepper et al., 2016). Although DNA methylation has been well studied in vertebrates, many other model organisms such as *Drosophila melanogaster, Caenorhabditis elegans*, and *Saccharomyces cereviceae* lack DNA methylation (Deaton and Bird, 2011). DNA methylation is

observed in <20% CpG dinucleotides in mouse embryonic stem (mES) cells and changes dynamically during differentiation (Meissner et al., 2008). Importantly, the majority of CpG islands at housekeeping and other gene promoters are usually unmethylated i.e. the promoters are permissive for transcription (Deaton and Bird, 2011; Meissner et al., 2008). DNA methylation is catalysed by DNA methyltransferases (DNMTs; "writer" enzymes), with DNMT1 maintaining DNA methylation through replication and DNMT3A/B catalysing *de novo* DNA methylation (Jackson et al., 2004). Interestingly, although the homozygous deletion of *DNMT1* in human ES (hES) cells causes rapid cell death, hES cells carrying a homozygous deletion of *DMNT3A* or *B* are still viable and pluripotent with only a progressive mild reduction in DNA methylation levels over many passages (Liao et al., 2015). DNA methylation is believed to cause transcriptional repression *via* the inhibition of TF binding and the recruitment of chromatin remodelling complexes (Deaton and Bird, 2011). In the following sections, I will describe polycomb group proteins (PcGs) and discuss, how their deposited marks are associated with gene repression. I will focus in particular on polycomb repressive complex 2 (PRC2) and the histone modification H3K27me3. I will also briefly discuss some of the literature regarding how the PRC1 and 2 complexes are recruited to target genes.

1.1.2.1 PRC1 and H2Aub

PcGs were first identified in *Drosophila melanogaster* by mutant screens searching for upstream regulators of the Hox genes. Mutations in PcGs led to aberrant activation of these genes and gross body patterning defects which showed their importance in maintaining correct developmental patterning (Lewis, 1978). Since then, two families of PcG complexes have been identified that are involved in transcriptional silencing; PRC1 and PRC2 (Morey and Helin, 2010; Simon and Kingston, 2009). Both of these complex families have many components that can combine in multiple ways to form complexes that do not always functionally overlap (Morey and Helin, 2010; Simon and Kingston, 2009).

Two components of PRC1 in humans, RING1B and BMI1, are required for the deposition of H2Aub (Cao et al. 2005). The H2Aub mark appears to be involved

in transcriptional repression *via* the inhibition of PolII elongation (Stock et al., 2007; Zhou et al., 2008). However, electron microscopy has shown that PRC1 components are capable of compacting nucleosomes associated with DNA, both in the presence and in the absence of histone tails *in vitro* (Francis et al., 2004). This evidence suggests that PRC1 mediates gene silencing through multiple pathways that do not only involve the H2Aub mark.

1.1.2.2 PRC2 and H3K27me3

The trimethylation of histone 3 at lysine 27 (H3K27me3) is deposited by PRC2. PRC2 complexes include EED (a "reader" enzyme), which binds to H3K27me3, and the histone methyltransferase EZH2 (a "writer" enzyme), which catalyses the H3K27me3 mark (Margueron et al., 2009). The importance of PRC2 in development is clear from early experiments where knocking out core PRC2 subunits genes *Eed*, *Ezh2* and *Suz12* proved to be embryonic lethal in mouse models (Faust et al., 1998; O' Carroll et al., 2001; Pasini et al., 2004). H3K27me3 can be removed by the histone demethylases ("eraser" enzymes), UTX (KDM2A), UTY and JMJD3 (KDM2B) (Agger et al., 2008; Walport et al., 2014).

The mechanisms by which PRC2 complexes confer transcriptional silencing are somewhat unclear. It is known that EZH2 interacts with enzymatically active DNMTs *in vitro* (Viré et al., 2006). However, PRC2 generally binds to CpG island promoters that are unmethylated (Ku et al., 2008). Despite this, EZH2 bound promoters in ES cells can sometimes gain DNA methylation during differentiation, thus preventing future transcription from the locus (Cedar and Bergman, 2009; Meissner et al., 2008; Mohn et al., 2008). It has also been shown that the H3K27me3 demethylase, JMJD3, is required for transcriptional elongation at genes enriched with this mark (Chen, et al. 2012; Estarás et al. 2013). This suggests that H3K27me3, like H2Aub, can prevent transcriptional elongation from occurring.

It is important to note that the recruitment of PRC2, and the resulting deposition of H3K27me3, occurs only *after* transcription has been turned off or chromatin compaction has happened (Hosogane et al., 2013; Yuan et al., 2012). It has

also been shown that loss of H3K27me3 in adult mouse intestinal cells does not necessarily result in activation of gene expression, and *vice versa* (Jadhav et al., 2016). This implies that the PRC2 mediated H3K27me3 mark is not required for gene repression (Hosogane et al., 2013; Jadhav et al., 2016; Yuan et al., 2012), although the evidence described above suggests it can play a direct role in gene silencing (Viré et al. 2006; Chen, et al. 2012; Estarás et al. 2013). A recent report in *Drosophila melanogaster* showed that local inheritance of the H3K27me3 mark, following excision of the closest polycomb repressive element (PRE), was sufficient for short-term inheritance of Hox gene repression (Coleman and Struhl, 2017). This suggests that H3K27me3, like H3K4me3, may be more indirectly involved in the maintenance of a transcriptional state during cell division and differentiation.

1.1.2.3 PcG recruitment

PcG complexes are recruited to specific DNA sequences called PREs in *Drosophila melanogaster* (Kassis and Brown, 2013). The mechanism of recruitment of PcG proteins in mammals has been more difficult to elucidate. In certain biological contexts it seems that PRC2 can be recruited by binding to specific TFs, such as the REST and SNAIL TF families, in neuronal progenitor cells (Arnold et al., 2013). There is also evidence that shows certain long non-coding RNAs are able to recruit PRC2. For example, PRC2 binds to the long non-coding RNA Xist through the JARID2 subunit, to facilitate the transcriptional silencing of one X chromosome in female mammals (Brockdorff, 2013; da Rocha et al., 2014). However, it appears that mammalian PRC2 can bind irrelevant RNA sequences (such as bacterial RNA) with the same efficiency as specific verified target RNA (such as the Xist long non-coding RNA) (Davidovich et al., 2015). These specific examples of recruitment do not account for the majority of PRC2 binding.

A prevailing theory proposes that PcG complexes, for the most part, 'sample' a large portion of genetic material in a sequence independent manner and then accumulate at 'target' sites, depending on the pre-existing chromatin and transcriptional context of that site (Klose et al., 2013). DNA methylation represses transcription and so it is thought that PcG recruitment is a method of preventing

transcription at genes that have escaped DNA methylation. Indeed, both PRC1 and 2 complexes colocalise with unmethylated CpG islands (Ku et al., 2008). It has also been shown that the introduction of ectopic unmethylated CpG islands is sufficient to recruit both PRC1 and PRC2 machinery in a cell (Lynch et al., 2011; Mendenhall et al., 2010). The recruitment of PRC2 to target loci in mES cells is partially interdependent with JARID2 recruitment (Li et al., 2010). JARID2 interacts with and modulates the activity of PRC2 (Li et al., 2010, Son et al., 2013). Initial studies suggested that this recruitment was due to the DNA-binding or nucleosome-binding activity of JARID2 (Li et al., 2010, Pasini et al., 2010; Peng et al., 2009). More recent work has shown that the interaction of JARID2 with lncRNAs is also involved in regulating PRC2 recruitment (Kaneko et al., 2014, da Rocha et al., 2014). A subset of mammalian PRC1 complexes can be recruited to unmethylated CpG islands *via* one of its subunits; the H3K4/K36 demethylase KDM2B (Farcas et al., 2012; Wu et al., 2013).

Once a PcG complex has been recruited to a CpG island it samples the local chromatin environment. The presence of histone marks associated with active transcription, H3K4me3, H3K27ac and H3K36me2/3, are all inhibitory to PRC2 enzymatic activity (Reynolds et al., 2012; Schmitges et al., 2011; Yuan et al., 2012). It is thought that PcG complexes identify correct target genes through uninhibited binding and then accumulate through positive feedback mechanisms (Klose et al., 2013). Indeed, the colocalisation of PRC1 and PRC2 components at silenced genes suggests that both complexes are involved in transcriptional repression at a subset of genes (Bracken et al., 2006). A number of reports have shown that the PRC2 mediated mark, H3K27me3, is required for PRC1 binding at certain loci (Cao et al., 2002; Wang et al., 2004; Zhen et al., 2016). Conversely, other groups have shown that the PRC1 mediated mark, H2Aub, is required for PRC2 recruitment at certain genes (Blackledge et al. 2014; Cooper et al. 2014). It would appear that a number of different mechanisms are involved in the accumulation of PcGs at individual loci.

1.1.3 Epigenetics, development and disease

Epigenetic mechanisms control many cellular decisions in stem cell biology, development, and disease. Global changes to the epigenetic landscape occur during early development. When an egg cell is fertilised and when primordial germ cells develop in the embryo, two large scale DNA demethylation and epigenetic reprogramming events occur, from which only certain heterochromatin and imprinted regions escape (Cantone and Fisher, 2013). As cells differentiate throughout development and adulthood, the epigenetic landscape continues to change, albeit at a more localised level. These dynamic changes can be observed by bisulphite sequencing (a method for determining the localisation of cytosine methylation) or ChIP-seq data sets from many different tissues, including pluripotent stem cells and lineage committed cells (Bernstein et al., 2010, 2012; Meissner et al., 2008; Mikkelsen et al., 2007). Many genetic studies have shown that knocking out or knocking down members of the DNMT, SET1-like and PcG complexes can cause disruptions in ES cell gene expression, differentiation, or can even be embryonic lethal (Carroll et al., 2001; Faust et al., 1998; Jackson et al., 2004; Liao et al., 2015; Pasini et al., 2004; Shilatifard, 2012). This highlights the importance of epigenetics as a marker for, and an instructor of, appropriate development. Furthermore, dysregulation of these pathways has been shown to cause disease in a number of cases (reviewed in Brazel & Vernimmen 2016). In this section I will briefly discuss some of the major genetic mutations in epigenetic regulators associated with cancer.

EZH2 is the most frequently mutated PRC2 component in cancer. Disruptions in H3K27me3 levels and localisation due to *EZH2* mutations have been identified in many solid tumours and blood malignancies (Morin et al., 2010; Souroullas et al., 2016). The DNA methyltransferase, DNMT3A, is also often catalytically inactivated by dominant-negative mutations in haematological malignancies (Ley et al., 2010; Yang et al., 2015). Indeed, the disruption of normal DNA methylation patterns in colorectal cancer cells has been correlated with a reduced expression of tumour suppressor genes and an overexpression of oncogenes (Irizarry et al., 2009). Disruption of epigenetic editing enzymes is commonly caused by chromosomal translocations, which can affect epigenetic regulators by creating novel fusion proteins with new functions. Chromosomal rearrangements involving the *MLL* gene

occur in over 70% of infant leukaemias, most of which cause a deletion of the catalytic H3K4 methlytransferase domain (Krivtsov and Armstrong, 2007). Despite the inactivation of the SET domain, the resulting fusion proteins cause overexpression of a number of different target genes, often *via* the recruitment of DOT1L, a H3K79 histone methyltransferase associated with transcriptional elongation (Bernt et al., 2011; Krivtsov et al., 2008). Mutations in the H3K27 demethylases JMJD3 and UTX have also been associated with malignancies (Arcipowski et al., 2016; Ntziachristos et al., 2014).

The contribution of epi-mutations to disease is less clear than that of genetic mutations in epigenetic regulators. It appears that specific DNA methylation and histone modification patterns can be used as biomarkers for diseases such as acute myeloid leukaemia (Figueroa et al., 2010), active ovarian cancer (Teschendorff et al., 2009) and colorectal cancer (Akhtar-Zaidi et al., 2012). These disease specific patterns are thought to play a role in the downregulation of tumour suppressor genes and upregulation of oncogenes, and so represent therapeutic targets. Epigenome-wide association studies (EWAS) are being used to identify differentially methylated regions associated with disease. Epi-mutations associated with Wilms tumour and Lynch syndrome, two solid malignancies, have been identified (reviewed in Hitchins 2015). The identification of epi-mutations is faced with fundamental challenges including intrinsic epigenetic variation between (and within) different cell types and developmental stages.

Epigenetic regulation of development is of paramount importance, as demonstrated by the vast numbers of diseases that occur when epigenetic pathways are disturbed. To more clearly understand the pathological mechanisms involved in these diseases we must first completely understand the exact function of their associated epigenetic marks in normal conditions. As described in the above sections, despite years of research the exact function of two individual histone marks, H3K4me3 and H3K27me3, in controlling transcriptional regulation remains ambiguous. Furthermore, these two marks are not mutually exclusive and may provide the cell with a novel set of instructions, different from each individual mark. In the following section I will discuss in detail the colocalisation of these two marks,

termed “bivalency”, describing how it is identified, established, and evidence regarding its function.

1.2 Epigenetic bivalency

I have previously discussed two histone modifications, H3K4me3 and H3K27me3, and their association with transcriptional activation and repression, respectively. At first glance it would seem that the individual presence of these histone marks would lead to two opposing transcriptional states, and thus the marks would themselves be mutually exclusive. However, this does not appear to be the case at a large number of genomic loci. Colocalisation of these two marks was first reported in ES cells (Azuara et al., 2006; Bernstein et al., 2006; Ku et al., 2008; Mikkelsen et al., 2007). The colocalisation of these two marks is termed “bivalency” and forms an important theme of this thesis. Other groups have referred to bivalency as the colocalisation of H3K4me2/H3K27me3 (Minoux et al., 2017) or H3K4me3/H2Aub (Brookes et al., 2012). Here, I define bivalency as the colocalisation of H3K4me3/H3K27me3, unless otherwise stated. In this section I will first discuss the methods used to study bivalency, and the resulting information these different assays provide. I will then describe where and how bivalent domains are established in a cell and conclude with a discussion of the functions associated with bivalency.

1.2.1 Methods used to study bivalency

Bivalency can be identified by comparing two individual ChIP-qPCR or ChIP-seq data sets that use antibodies against H3K4me3 and H3K27me3 on chromatin preparations from the same population of cells. However, there is a chance that in an apparently homogenous cell population, for example a mES cell line, the occurrence of an epigenetically mixed population (Figure 1.1B) can lead to misinterpretation of data. Classic ChIP experiments require thousands to millions of cells (a classic ChIP protocol is used in this thesis and described in Chapter 2: Materials and Methods pp. 55). Even protocols optimized for cross-linked or native

ChIP using ultra-low cell numbers require at least 500-1,000 cells (Brind'Amour et al., 2015; Zheng et al., 2015). There is a significant risk that these cell populations contain some degree of heterogeneity. This heterogeneity can occur as a result of

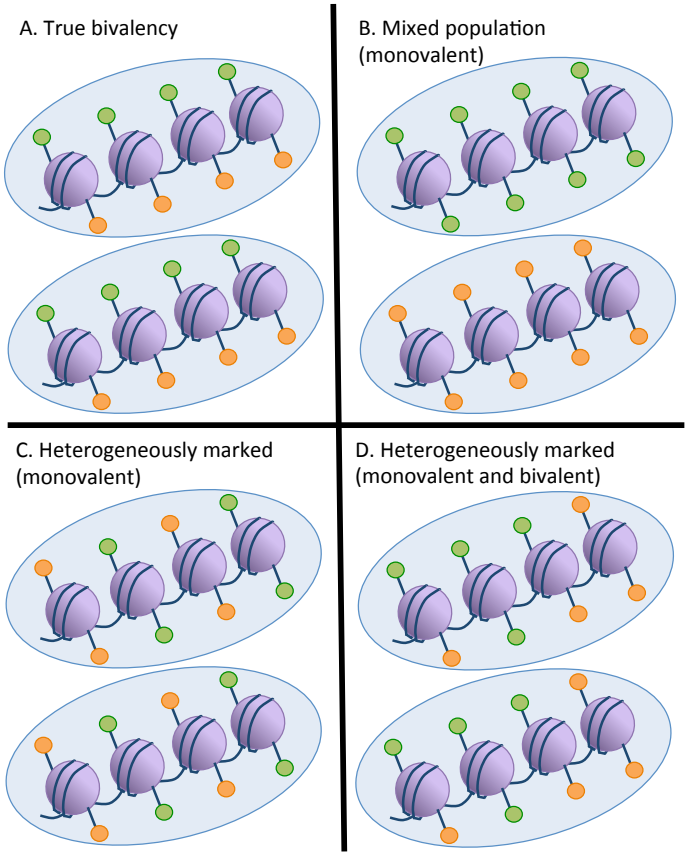


Figure 1.1 Different epigenetic landscapes can appear bivalent using classical ChIP protocols. A model of the possible epigenetic landscapes at a TSS that appears bivalent using classical ChIP protocols. The cells (blue ellipses) contain individual nucleosomes (purple circles) that carry the histone modifications H3K4me3 (green circles) or H3K27me3 (orange circles) on different histone tails.
A. When H4K4me3/H3K27me3 colocalise on the same nucleosome, the promoter can be called truly bivalent.
B. A promoter can be monovalently marked with either H3K4me3 or H3K27me3 in different cells. This is an epigenetically mixed population.
C. Neighbouring nucleosomes can be monovalently marked with either H3K4me3 or H3K27me3 generating a heterogeneously marked promoter.
D. Neighbouring nucleosomes can be monovalently marked with either H3K4me3 or H3K27me3, or be bivalently marked with both histone modifications, generating a heterogeneously marked promoter.

mixture of different cell types. For example, whole tissue preparations (e.g. whole embryos) contain many different cell types (e.g. cells from the endodermal,

mesodermal and ectodermal germ layers). Heterogeneity can also occur as a result of individual cells of the same cell type (e.g. ES cells) containing different histone modification at the same locus (Figure 1.1B). Moreover, this admixture of cells, containing either repressive or active chromatin marks at certain genes, may masquerade as true bivalent domains in ChIPs on large populations of cells (Figure 1.1).

A number of methods have been developed which can be used to elucidate whether the H3K4me3 and H3K27me3 marks are truly colocalised and not observed as a result of an epigenetically mixed or heterogeneous population. Sequential ChIP (re-ChIP) is a method whereby chromatin is immunoprecipitated with an antibody with high affinity towards one protein and then subjected to a second round of immunoprecipitation (IP) with an antibody that recognises a second protein (Hatzis and Talianidis, 2002). The aim is to recover only the chromatin that simultaneously carries both proteins. Re-ChIP studies using antibodies that identify H3K4me3 and H3K27me3 marks support the existence of bivalency at some loci (Bernstein et al., 2006; De Gobbi et al., 2011; Ku et al., 2008; Mikkelsen et al., 2007). However, re-ChIP studies are limited in their ability to identify loci with a low enrichment of either mark as the quantity of material is severely restricted due to repeated rounds of IP.

Mass spectrometry analysis has been used on mononucleosomes that were immunoprecipitated with antibodies that recognise histone modifications to determine if these marks occur on the same nucleosome (Voigt et al., 2012). This study confirmed that H3K27me3 and H3K4me3 can colocalise to the same nucleosome and that these marks occupy opposite histone tails when found in a single nucleosome (Voigt et al., 2012). However, mass spectrometry analysis does not provide information regarding the genomic localisation of the bivalently marked nucleosome. It is therefore conceivable that although certain nucleosomes may be bivalently marked, some bivalency observed in classical ChIP studies may be due to epigenetically heterogeneous cell populations.

In the last year, there have been a number of breakthroughs in the identification and mapping of bivalent marks on single molecules (Sadeh et al., 2016;

Shema et al., 2016; Weiner et al., 2016). Two groups used combinatorial indexed ChIP (co-ChIP-seq), a modified version of re-ChIP, to identify colocalisation of histone marks in cells from yeast and mice (Sadeh et al., 2016; Weiner et al., 2016). Co-ChIP-seq involves the ligation of barcoded adaptors during the first round of IP. This barcoding allows the samples from different IPs to be pooled before the second round of IP. This prevents the loss of information of which DNA fragments were immunoprecipitated with the first antibody, and prevents the loss of experimental material. Using this method, Weiner *et al.* (2016) identified >17,000 bivalent regions across the genome in mES cells. Another group recently developed single molecule imaging and sequencing to identify bivalent regions in mES cells (Shema et al., 2016). In this study, chromatin fragments were ligated to fluorescent adaptors and captured on slides. These slides were then incubated with fluorescently labelled antibodies with high affinity for various histone modifications and the dynamic binding of these antibodies was recorded. Finally, proteins were removed from the DNA bound to the slide, and single molecule sequencing was performed to map the genomic localisation of the histone modifications. Shema *et al.* (2016) also used this method on individual histones isolated from mES cells and confirmed that the vast majority of putative bivalently marked nucleosomes carry only one of these marks (H3K4me3 or H3K27me3) on each histone tail (Voigt et al., 2012).

These technical advancements have confirmed that individual nucleosomes can indeed carry both the H3K4me3 and H3K27me3 mark (Voigt et al., 2012) and have managed to map this bivalency to genomic loci in a number of cell types (Shema et al., 2016; Weiner et al., 2016). In the following section I will discuss the positioning of bivalent nucleosomes within a promoter and the types of cells and genes bivalency is observed in.

1.2.2 Occurrence of bivalency

Highly organised nucleosome positioning is observed during zebrafish embryogenesis in bivalent and H3K4me3 monovalent promoters, but not in H3K27me3 monovalent promoters (Zhang et al., 2014). Within bivalent promoters,

re-ChIP-seq analysis has shown that bivalent mono-nucleosomes tend to immediately flank the TSS (Sen et al., 2016). Sen *et al.* (2016) used classical ChIP-seq in the NCCIT human pluripotent germ cell tumour derived cell line, to show that broad domains of H3K27me3 (~4-8 kb) can overlap with narrower peaks of H3K4me3 (~1-4 kb) slightly upstream of the TSS. However, the region of true bivalent mononucleosomes, as determined by re-ChIP-seq analysis, is even narrower (0.5-2 kb) and tends to occur in the immediate vicinity of the TSS (Sen et al., 2016). This analysis revealed that although bivalent mononucleosomes tend to localise at the TSS, bivalent mononucleosomes can also occur throughout the promoter region, alongside monovalent mononucleosomes (Figure 1.1D; Sen et al. 2016).

Bivalency was first identified in pluripotent ES cells from humans and mice (Azuara et al., 2006; Bernstein et al., 2006; Ku et al., 2008; Mikkelsen et al., 2007). Since then, many data sets have been produced using classical ChIP methods assessing individual marks in hES and mES cell populations. Two groups integrated and analysed 11 published datasets to determine a list of high confidence bivalent promoters (Court and Arnaud, 2017; Mantsoki et al., 2015). The previously described study by Weiner *et al.* (2016) identified >17,000 bivalent domains in hES cells in 1 kb sliding windows across the entire genome. Here, a set of ~4,300-5,000 high confidence bivalent promoters, within 3-4 kb of an annotated TSS, were identified in hES cells (Court and Arnaud, 2017; Mantsoki et al., 2015). These high confidence bivalent promoters tend to be unmethylated CpG islands and tend to be associated with genes that have low expression in ES cells and tissue-specific expression patterns after differentiation (Court and Arnaud, 2017; Mantsoki et al., 2015). Importantly, over 50% of bivalency is conserved between promoters from mES and hES cells, suggesting an important functional role of bivalency (Mantsoki et al., 2015).

Bivalency is not restricted to pluripotent cells. A number of reports have identified bivalency at a large number of gene promoters in differentiated cells, for example lineage committed cell lines derived from mES cells (Mikkelsen et al., 2007), primary human CD34+ cord blood cells (Lorzadeh et al., 2016), primary human CD4+ memory T cells (Kinkley et al., 2016) and adult mouse intestinal cells

(Jadhav et al., 2016). These reports have often shown that a subset of bivalent domains overlap between ES cells and differentiated cells, but that certain bivalent domains are lost or new bivalent domains are formed. For example, Weiner *et al.* (2016) showed that of the 23,167 bivalent domains discovered in total in 5 different tissues, 5,786 change dynamically between mES cells and mouse adult brain, kidney, liver and lung tissues (from genome wide co-ChIP-seq analysis not restricted to promoter regions). Bivalency occurs in many different pluripotent and differentiated cell types and changes dynamically during differentiation. In the following section I will discuss how a bivalent state is formed at promoters.

1.2.3 Establishment of bivalency

A number of factors appear to be important in the establishment of bivalency. Firstly, an unmethylated CpG island is required to recruit members of the PcG and SET1 or SET1-like complexes that catalyse H3K27me3 and H3K4me3 (see section 1.1.1, pp. 2 and section 1.1.2, pp. 5). Once these complexes have been recruited, no single complex can be allowed to outcompete the other and establish a monovalent domain. A balance between the two marks must be established by means of appropriate deposition and removal of the marks. Below, I will discuss the mechanisms involved in the establishment of bivalency.

It has been well reported that bivalency predominately occurs in promoters containing unmethylated CpG islands (Court and Arnaud, 2017; Meissner et al., 2008, Mantsoki et al., 2015). Indeed, it has recently been shown that DNA demethylation by the TET family is important for both the establishment and maintenance of bivalency (Kong et al., 2016). Overexpression of the DNA demethylase genes, *TET1/2*, in HEK293T cells caused DNA demethylation at a subset of hypermethylated CpG islands promoters and this was sufficient for the *de novo* establishment of bivalent domains (Kong et al., 2016). It has also been shown that demethylation of intragenic CpG (iCpG) islands (within the gene-body), is important in regulating the establishment of bivalent domains (Lee et al., 2017). Hypermethylation of iCpG islands is usually required for high activation of gene

expression but cell type specific methylation of iCpG islands can cause loss of bivalency at the associated gene promoter (Lee et al., 2017).

It is possible that the deposition of H3K4me3 occurs before H3K27me3 at bivalent promoters. MLL2 is required to establish the H3K4me3 mark at bivalent promoters in mES cells (Denissov et al., 2014; Hu et al., 2013). While MLL1 is not required for the establishment of H3K4me3 at bivalent promoters (Denissov et al., 2014), it is required to establish H3K4me2 at a subset of H3K27me3 marked CpG island promoters (Rickels et al., 2016). Dynamic changes have been observed in the presence of H3K4me3 at bivalent genes during the cell cycle and it has been shown that the MLL1/2 proteins are recruited to bivalent genes in a cell cycle dependent manner (Grandy et al., 2015).

Once a H3K4me3 domain has been established at an unmethylated CpG island promoter, PRC2 mediated deposition of H3K27me3 is required to establish a bivalent domain. It had previously been shown that the presence of H3K4me3 can inhibit PRC2 methylation of H3 in nucleosomes (Schmitges et al., 2011). However, this inhibition of PRC2 does not extend to the 'free' unmarked H3 tail in asymmetrically H3K4me3 marked nucleosomes (Voigt et al., 2012). It has thus been proposed that when the K4me3 mark is removed from at least one H3 copy in a nucleosome by a histone demethylase, it can then be methylated by PRC2 resulting in bivalently marked domains (Voigt et al., 2013). Indeed, it has since been shown that the H3K4me2/3 demethylase, JARID1B (KDM5B), colocalises with the majority of bivalent genes in mES cells (Kidder et al., 2014). In the absence of *Jarid1b*, bivalent gene expression is delayed or reduced during differentiation of mES cells, suggesting that JARID1B is important in bivalent gene regulation (Kidder et al., 2013). Another H3K4me2/3 demethylase, KDM2A, is also recruited to bivalent promoters (Grandy et al., 2015), indicating that these histone demethylases may perform redundant functions at bivalent genes.

1.2.4 Functions of bivalency

Once a bivalent domain has been established at a gene promoter, these genes are expressed at low or undetectable levels with a high degree of variation in expression between single cells (Bernstein et al., 2006, Court and Arnaud, 2017; Mantsoki et al., 2016, Mikkelsen et al., 2007). It has been proposed that bivalency poises developmental genes for timely activation or complete repression (Bernstein et al., 2006). It is thought that this activation or repression occurs following the resolution of bivalency, i.e. the conversion to a monovalent state. Bivalency resolution readily occurs during differentiation. Mouse adult brain tissues have lost bivalency at nearly 5,000 regions compared to ES cells (Weiner et al., 2016). To explore the function of bivalency during development, a number of studies have achieved the global depletion of one of the bivalent histone marks and analysed the resulting phenotypes. In this section I will discuss the results of these experiments and the various models of bivalency function.

It has been proposed that the existence of the H3K4me3 mark on bivalent genes is important for their timely activation following differentiation. However, the loss of H3K4me3 from bivalently marked genes in mES cells does not appear to affect the expression of bivalent genes following differentiation, or the differentiation capacity of the cells (Denissov et al., 2014; Hu et al., 2013). Furthermore, the loss of the H3K4me2 mark at H3K4me2/H3K27me3 marked promoters only causes expression changes in <200 genes bound by MLL1 at the TSS (Rickels et al., 2016). These results do not suggest that the H3K4me2/3 mark at bivalent genes is important for their activation after differentiation.

Another study focused on analysing gene expression changes after the depletion of the H3K27me3 mark on bivalent promoters (Jadhav et al., 2016). At a bivalent promoter, the loss of the H3K27me3 mark, and conversion to a monovalent H3K4me3 marked state, is proposed to mediate gene activation following differentiation. Indeed, it has been shown that the H3K27me3 demethylase, UTX, is required for activation of bivalent genes after the induction of mES cell differentiation (Dhar et al., 2016). Jadhav *et al.* (2016) inactivated PRC2 *via* an inducible Cre-mediated knockout of *Eed* in adult mouse intestinal cells, which led to

epithelial defects in the intestinal crypts. The knockout of *Eed* caused a loss of H3K27me3 and an increase in gene expression from genes that were specifically marked with H3K27me3 in the intestine. The increase in gene expression was directly proportional to the basal levels of H3K4me2/3 that were present at the promoter. This suggests that the tissue-specific localisation of H3K27me3 is important in maintaining repression of bivalent genes in these cells. Furthermore, the level of H3K4me2/3 at a bivalent promoter may be important in tuning gene expression levels following the removal of H3K27me3 (Jadhav et al., 2016).

It is also possible that the recognition of bivalently marked chromatin by specific factors is developmentally important. Notably, a transcriptional regulator UTF1 is enriched at bivalent genes in mES cells and appears to compete with PRC2 for binding sites, thus limiting PRC2 loading and H3K27me3 deposition (Jia et al., 2012). Conversely, UTF1 also promotes the degradation of mRNAs transcribed from bivalent genes *via* the recruitment of the mRNA decapping enzyme DCP1A (Jia et al., 2012). A motif has been identified that is enriched in high confidence bivalent promoters suggesting that other modifiers of bivalent genes that bind to this sequence may yet be identified (Mantsoki et al., 2015).

From these studies, a universal function of bivalency in transcriptional regulation during development is unclear. The classic model that bivalency is required for timely gene activation or complete repression following differentiation and resolution does not fit with the evidence presented to date. An alternative model suggests that MLL2 binding at bivalent genes prevents the binding of SET1C, which is only found at actively transcribing H3K4me3 marked promoters (Denissov et al., 2014). Perhaps these genes are only activated once bound by SET1C. It seems that bivalency may be important at a tissue-specific level for the repression of genes. If this bivalency were resolved in these cells *via* H3K27me3 removal, the levels of H3K4me2/3 would determine the level of gene induction (Jadhav et al., 2016). The specific manipulation of these marks at particular bivalent genes and analysis of dynamic molecular changes during cell differentiation would help elucidate the role of bivalency in gene regulation. Another chromatin feature that is associated with regulating gene expression during development are long-distance regulatory elements

(referred to here as enhancers). In the following section I will discuss enhancers and focus on the models that describe their actions and functions.

1.3 Enhancers

A number of proximal and distal regulatory regions are involved in controlling gene transcription. Promoters are regions of DNA that are required for transcriptional activation at a gene TSS (Juven-Gershon and Kadonaga, 2010). Enhancers are DNA sequences that, when activated in a tissue-specific manner, can increase the probability their target gene is transcribed (Fiering et al., 2000). Enhancers can be located long distances of up to 1 Mb from their target promoters (Lettice et al., 2003). In this section I will discuss how enhancers are identified and how they function to activate the transcription of a target gene.

1.3.1 Identifying enhancers of transcription

In 1981 the first enhancer was isolated, consisting of 72 bp repeats from Simian virus-40 (Banerji et al., 1981). This short regulatory region was found to increase the transcription of a β-globin gene when cloned into a reporter plasmid and transfected into HeLa cells (Banerji et al., 1981). The basic principles of enhancer validation studies have remained largely similar since that time. Predicted enhancers are often cloned into a reporter vector with a minimal promoter and introduced into cell lines or developing embryos. The enhancer activity can then be visualised and spatio-temporal regulation can be observed (Pennacchio et al., 2006). However, the temporally regulated activity of enhancers can sometimes make them difficult to validate, as they may not be active at the developmental stage they are tested. The site of random integration of the reporter construct may also reduce the apparent enhancer activity. Position effect variegation can cause up to 1,000 fold variation of enhancer activity based on the surrounding chromatin architecture, genomic context or transcriptional activity at the site of integration (Akhtar et al., 2013).

Over a decade ago, enhancers began being identified by comparative genomics; non-coding regions undergoing evolutionary constraint were often found

to elicit enhancer activity when tested, but many false positives indicated that browsing DNA sequence alone would not be sufficient for enhancer identification (Pennacchio et al., 2013). Since then a number of studies have identified a unique combination of chromatin features that define enhancers. Unlike promoters, enhancers tend to be located in CpG poor regions in mammals (Andersson et al., 2014a). Active enhancers are located in open chromatin regions that are depleted of nucleosomes and so, like other regulatory and coding regions, can be identified by hypersensitivity to the non-specific endonuclease DNaseI (Dorschner et al., 2004). A specific histone modification signature has also been associated with enhancers. An enrichment of H3K4me1 and depletion of H3K4me3 marks enhancer regions, while the inverse is indicative of a promoter region (Heintzman et al., 2007, 2009). The H3K27ac mark, deposited by the histone acetyltransferase and transcriptional coactivator p300, is enriched at both active enhancers and promoters (Creyghton et al., 2010; Visel et al., 2009). Bidirectional transcription of capped enhancer RNA (eRNA) also defines active enhancers (Andersson et al., 2014a, 2014b). Generally, enhancers can be defined as being within CpG poor, nucleosome depleted regions that are enriched with H3K4me1 and depleted of H3K4me3. Furthermore, active enhancers can be identified by the enrichment of H3K27ac, p300 and by the production of bidirectional capped transcripts. Multiple reporter assays have shown that these marks are very likely to identify functional enhancers when tested (Andersson et al., 2014a; Heintzman et al., 2007). Enhancers have also been identified that carry a combination of marks (H3K27ac and DNA methylation) associated with both active and inactive enhancers (Charlet et al., 2016).

1.3.2 Functions of enhancers

Once enhancers have been identified it is then important to understand how they function. In this section I will discuss the various models of how enhancers regulate gene activity and briefly describe the importance of TF binding to enhancers in mediating this activity.

1.3.2.1 A stochastic model for enhancer activity

The vast majority of enhancers are only activate in particular tissues and at specific developmental time points (Arner et al., 2015; Nord et al., 2013). However, enhancer activation does not only occur in response to developmental cues. Challenge experiments have shown that some "latent" enhancers are only activated in terminally differentiated cells in response to an external stimulus (Ostuni et al., 2013). Recent data from the FANTOM Consortium has suggested that enhancers can rapidly become activated in response to a change in cell state, and that this activation occurs before changes in target gene transcription (Arner et al., 2015). Once enhancers have become activated they can regulate the expression of their target genes (Arner et al., 2015).

Two models of enhancer action have been proposed (reviewed in Fiering et al. 2000). Originally, a "rheostatic" model proposed that enhancers increase the rate of transcription at a target gene. However, much evidence instead supports the second "stochastic" model which proposes that enhancers increase the probability of transcription occurring at a target gene (Bartman et al., 2016; Fukaya et al., 2016; De Gobbi et al., 2017). Individual enhancers can activate target gene expression (Banerji et al., 1981), or more complex clusters of enhancers such as "super enhancers" (Hay et al., 2016; Hnisz et al., 2013), originally termed "locus control regions" (Grosveld et al., 1987; Talbot et al., 1989), can work together to control the expression of a gene. In the following section I will discuss the various mechanisms that have been suggested to explain how enhancers regulate gene activation.

1.3.2.2 Mechanisms controlling enhancer-mediated activation

A large amount of studies, often using fluorescent *in situ* hybridization (FISH) visualization or 3C based technology, have shown that enhancers physically interact with their target promoters through the formation of a chromosomal loop (Amano et al., 2009; Dekker et al., 2002; Kieffer-Kwon et al., 2013). This physical interaction coincides with an increased concentration of factors required for transcriptional activation at the target promoter, including PolII (Vernimmen and Bickmore, 2015; Vernimmen et al., 2007, 2011). Of interest, the looping structure itself has been

shown to be involved in the activation of target gene expression in the absence of TFs that are usually essential for transcription to occur (Deng et al., 2012).

The binding of tissue-specific TFs to enhancers is integral in mediating spatiotemporal regulation of the target gene expression (Spitz and Furlong, 2012). Recently, a number of elegant experiments from the Levine lab (Princeton University, New Jersey) showed the fundamental importance of the order, orientation, spacing (together termed "syntax") and sub-optimisation of TF binding sites in enhancers (Farley et al., 2015, 2016). Farley *et al.* (2015) randomly mutated TF binding sites and TF binding site syntax in enhancer sequences and generated reporter libraries with these mutations. These mutant libraries were then electroporated into *Ciona intestinalis* embryos and screened for changes in reporter activity relative to the unmutated enhancer (Farley et al., 2015). It was found that the combination of optimal TF binding sites and optimal syntax boosted reporter expression, but also abolished enhancer-mediated tissue-specific reporter activity (Farley et al., 2015, 2016). The presence of non-canonical TF binding sites and sub-optimal syntax appears to be extremely important in conferring robust tissue-specific expression (Farley et al., 2015, 2016).

It has also been shown that a small number of single-nucleotide polymorphisms (SNPs) in enhancers can result in dramatic morphological changes (Frankel et al., 2011; Lettice et al., 2003). SNPs in the *Drosophila sechellia* and *Drosophila melanogaster* enhancers are important in the tissue-specific expression of the target TF gene, *shavenbaby*, and result in dramatic morphological changes (Frankel et al., 2011). More recently it has also been shown that these SNPs formed a novel TF binding site for a potent repressor in the *Drosophila sechellia* enhancer that is important in limiting ectopic expression of *shavenbaby* (Preger-Ben Noon et al., 2016). A study that identified active enhancers (H3K27ac+/H3K4me3-) and active promoters (H3K27ac+/H3K4me3+) and compared their underlying genomic sequences in liver tissue across 20 mammalian species showed that the enhancers are evolving rapidly, at approximately three times the rate of promoters (Villar et al., 2015). It seems that the subtle changes in TF binding sites in enhancers can be extremely important in changing the spatiotemporal regulation of their target genes,

and that this may be an extremely important mechanism in evolution. In the following section I will discuss how enhancer SNPs, and also enhancer deletions, have been associated with a number of genetic diseases in humans.

1.3.2.3 Enhancers and disease

Point mutations (or SNPs) in enhancers have been shown to contribute to a number of diseases, indicating their importance in regulating gene transcription during development (reviewed in Brazel & Vernimmen 2016). As described in the above section, SNPs can affect enhancer activity by changing TF binding or syntax (Farley et al., 2015, 2016). A number of SNPs in enhancers have been identified that are associated with disease. For example, Hirschsprung disease, a multigenic, heritable disorder affecting the ganglion cells in the large intestine or gastrointestinal tract, is associated with SNPs in the enhancers of *RET* and *SOX10* (Bondurand et al., 2012; Emison et al., 2005; Lecerf et al., 2013). Point mutations in the enhancer of *Sonic Hedgehog* (*SHH*) have been shown to cause preaxial polydactyly (formation of extra digits) in humans (Lettice et al., 2008, 2003). A number of independent studies have identified distinct solid tumour risk associated SNPs in the 8q24 enhancer region (Ahmadiyeh et al., 2010). This enhancer region has a number of different target genes, including the oncogene *MYC* (Cai et al., 2016). Some of these enhancer SNPs have been shown to modify TF binding affinity. For example, a prostate cancer risk allele enhances interactions with the TF forkhead box protein A1 (FoxA1), while a renal cancer risk allele modulates binding of the hypoxia-inducible TFs (Grampp et al., 2016; Jia et al., 2009).

The deletion of enhancers can also cause genetic diseases. For example, the deletion of the enhancers of globin genes has been shown to cause α- and β-thalassaemia in certain patients *via* down regulation of their target genes (Vernimmen, 2014). The characterization of the molecular basis of β-thalassaemias in patients led to the identification of regulatory elements controlling the β-globin locus (Kioussis et al., 1983). The existence of an enhancer region for the α-globin cluster in humans was first identified due to a large 62 kb deletion upstream of the α-globin cluster that caused α-thalassemia in a patient despite no mutations in their coding gene or promoter sequences (Hatton et al., 1990). Since then, a number of

causative deletions of varying size spanning the α-globin regulatory region have been identified from patients with α-thalassemia (reviewed in Higgs 2013). Many advances in our understanding of transcriptional regulation were made by studying the globin loci. In the following section I will describe the α-globin locus and focus on the role of enhancers in regulating α-globin transcription.

1.4 Transcriptional regulation at the α-globin locus

The α- and β-globin loci are two of the most well studied models of transcriptional regulation. In fact, a β-globin gene from rabbit was the first mammalian gene ever isolated with an intact promoter region (Maniatis et al., 1978). Haemoglobin is composed of α- and β- globin subunits and is expressed in the red blood cells of nearly all vertebrates and some invertebrates (McGrath and Palis, 2008). Haemoglobin functions as an oxygen transporter, ensuring that oxygen is delivered from the lungs to all other tissues in the body through the circulatory system. Different forms of haemoglobin are produced at the embryonic, foetal and adult stages of development from different tissues and the 'switching' of expression between different globins has been studied extensively (reviewed in Dzierzak & Philipsen 2013). In adults, the haematopoietic stem cells that reside in the bone marrow can differentiate into common myeloid progenitor cells which give rise to the erythroid lineage (Higgs et al., 2008). Throughout erythropoiesis a cascade of TFs progressively accumulate at the globin promoters and associated enhancers during lineage restriction, culminating in the eventual maximal globin expression in intermediate erythroblasts (Ferreira et al., 2005; Higgs and Wood, 2008; Philipsen and Hardison, 2017). In this section, I will briefly describe the α-globin locus and discuss in detail a humanised mouse model used to study α-globin gene regulation.

1.4.1 Organisation of the α-globin locus

The α-globin locus is well conserved through vertebrate evolution. A region of conserved synteny has been identified across ~135 kb of the α-globin locus that spans ~500 million years of evolution in 22 species (Hughes et al., 2005; Philipsen and

Hardison, 2017). In humans, the α-globin locus is located on Chromosome 16 and the five functional α-globin genes are arranged in the order shown in Figure 1.2

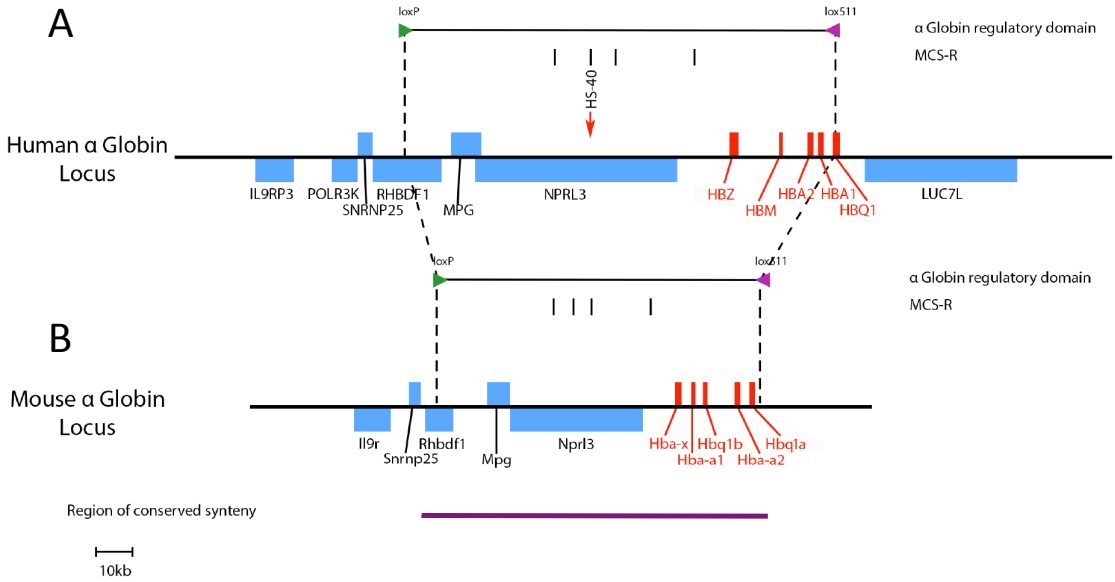


Figure 1.2 Chromosomal organisation of the (A) human and (B) mouse α-globin loci. The five functional α-globin and α-globin-like genes are shown as labelled red boxes. The human genes include *HBZ* (ζ-globin; expressed in embryonic stages), *HBM* (μ-globin; high mRNA expression, no protein expression detected), *HBA1* and *2* (α-globin; expressed in adult stages) and *HBQ1* (θ-globin; low expression in foetal stages). The mouse genes include *Hba-x* (ζ-globin; expressed in embryonic stages), *Hba-a1* and *2* (α-globin; expressed in adult stages) and *Hbq-1a* and *b* (θ-globin; lowly expressed in foetal stages). Other mouse and human genes in the locus are shown as labelled blue boxes. Genes shown above the line are transcribed from the forward DNA strand, while genes shown below the line are transcribed from the reverse strand. A region of conserved synteny is shown in purple and MCS-R (black vertical lines) refers to previously annotated multispecies conserved regulatory regions. In humanised mice, the region above the human locus, delimited by dashed lines, was exchanged from a human bacterial artificial chromosome (BAC) to mES cells. In the process of this exchange, one copy of the region above the mouse locus, delimited by dashed lines, was deleted. The exchanged regions are flanked by a 5` *loxP* (green triangle) and a 3` *lox511* (purple triangle) site. This figure is adapted from Wallace *et al.* (2007).

(Hughes et al., 2005; McGrath and Palis, 2008). A number of multispecies conserved regions have been identified across the locus (Hughes et al., 2005). To study the

regulation of human α-globin transcription *in vivo,* Wallace *et al.* (2007) generated a 'humanised' mouse, in which the ~85 kb mouse α-globin cluster was replaced with ~117 kb of the orthologous human α-globin cluster (Figure 1.2). The inserted region included all the human α-globin regulatory elements, allowing human α-globin transcriptional regulation to be studied *in vivo* (Wallace et al., 2007). A crucial difference between the human and mouse α-globin loci is the presence of CpG islands at the human α-globin genes that are absent in the mouse (Lynch et al., 2011). This is reflected by the fact that the human α-globin locus is bivalent in hES cells while the mouse α-globin is not bivalent in mES cells (De Gobbi et al., 2011). In the following section, I will discuss the roles of remote regulatory regions in the control of α-globin gene expression and in particular focus on data from the humanised mouse model.

1.4.2 Roles of remote regulatory elements

The long range regulatory elements within the α-globin locus have been well characterised in both mice and humans (Higgs et al., 2008). The identification of patients with α-thalassemia caused by deletions of α-globin enhancers, has greatly contributed to the analysis of α-globin transcriptional control (Hatton et al., 1990; Higgs, 2013; Higgs et al., 2008). Genetically engineered enhancer deletions in humanised and wild type mice have also been studied to elucidate the contribution of these enhancers to α-globin transcriptional regulation (Hay et al., 2016; Wallace et al., 2007). Notably, a recent study has shown that only fractional differences in mouse α-globin RNA production are observed when individual enhancers in the locus are deleted, resulting in no change to the overall haemoglobin levels (Hay et al., 2016).

In contrast, a much more dramatic phenotype is observed following the deletion of the human HS-40 enhancer (Wallace et al. 2007; Vernimmen et al. 2009). The HS-40 enhancer in the human α-globin locus is ~60 kb away from the α-globin genes (Figure 1.2A). To probe the contribution of the HS-40 enhancer, a humanised mouse with a HS-40 deletion was generated (Wallace et al., 2007). Removal of the HS-40 enhancer caused a severe reduction in α-globin expression to <2% of normal, indicating its critical role in α-globin regulation (Wallace et al. 2007; Vernimmen et

al. 2009). Indeed, many lines of evidence have shown that HS-40 is the only regulatory domain in the human locus that is capable of enhancing α-globin on its own (reviewed in Higgs & Wood 2008). A small number of cases of α-thalassaemia have been identified that are caused by homozygous deletions of the HS-40 enhancer in patients that have intact α-globin genes (Coelho et al., 2010; Sollaino et al., 2010; Wu et al., 2016). These patients present with moderate to severe forms of α-thalassaemia consistent with the deletion of three of the four α-globin genes, however this does not fully equate to the severity of α-globin dysregulation observed in the humanised mouse model with a HS-40 deletion.

Detailed molecular comparisons were made of the α-globin locus in erythroid cells from the wild-type humanised mouse (WTH) and the HS-40 deleted humanised mouse (ΔHS-40; Figure 1.2; Vernimmen et al. 2011; De Gobbi et al. 2017). Immunofluorescence and single cell reverse-transcriptase (RT-) PCR showed that only ~50% of terminally differentiated erythroid cells from ΔHS-40 mice express detectable levels of human α-globin (Figure 1.3A, B). Furthermore, uncharacteristic bivalency is observed at the α-globin genes in erythroid cells from ΔHS-40 mice (Figure 1.3C). It is possible that an underlying epigenetically mixed population of cells is reflected by the heterogeneous phenotype that is observed in the ΔHS-40 cells. If this hypothesis were true, the α-globin expressing ΔHS-40 cells would feature H3K4me3 at the α-globin genes while the non-expressing ΔHS-40 cells would instead feature H3K27me3 at the α-globin genes. Transgenic experiments could be used to generate humanised α-globin reporter mice to test this theory. In the following section, I will discuss a number of transgenic technologies, including the recombineering technologies that were used to generate these humanised mouse models. I will also discuss single cell reporter systems that can be used to track gene expression in individual cells.

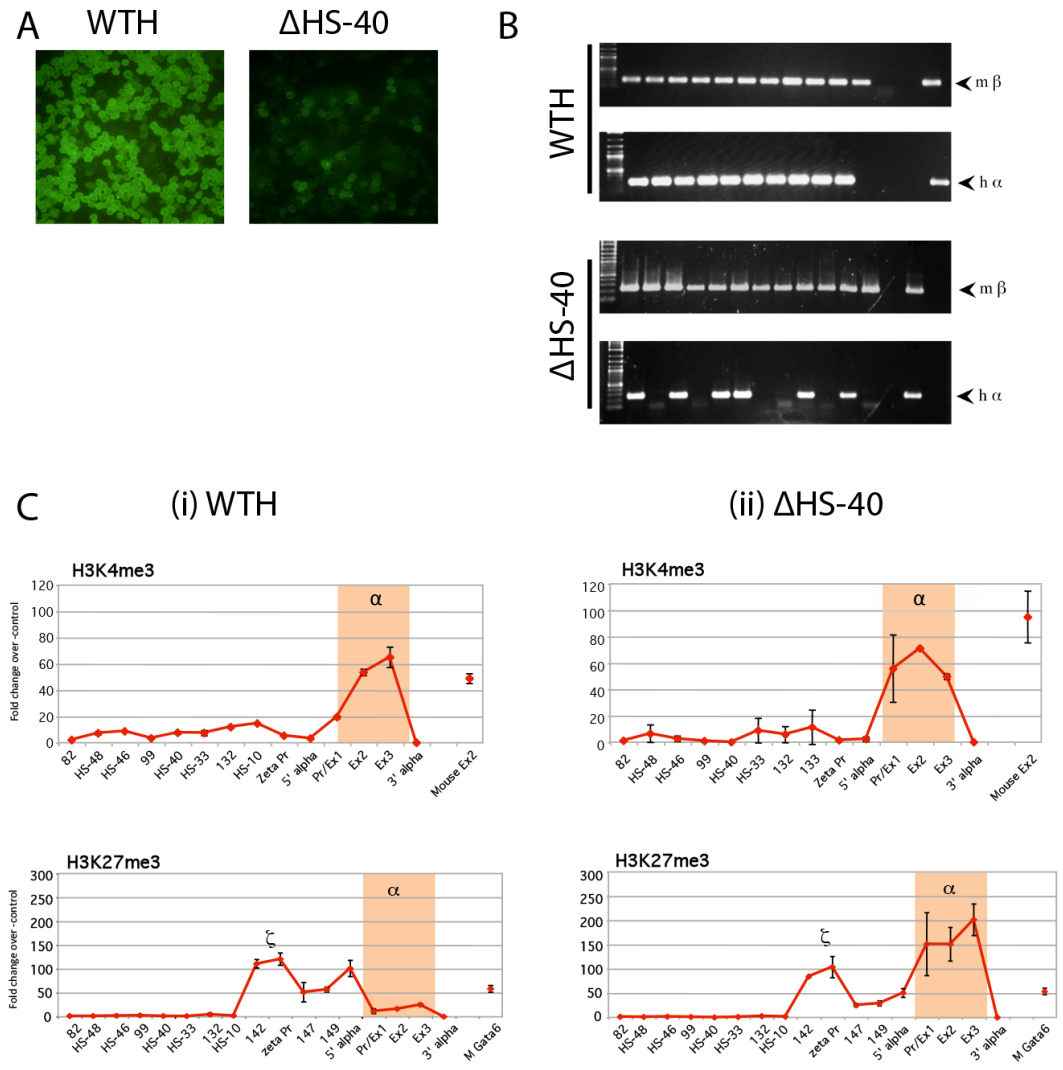


Figure 1.3 Heterogeneous α-globin expression patterns in ΔHS-40 humanised mouse erythroid cells.

A. Immunofluorescence staining of WTH and ΔHS-40 mouse humanised red cells using an antibody specific for human α-globin.

B. Single cell RT-PCR analysis in WTH and ΔHS-40 humanised erythroid cells. Each lane corresponds to the same single cell. The expression of human α-globin (hα) has been compared to mouse β-globin (mβ).

C. ChIP-qPCR results using the H3K4me3 (top) and H3K27me3 (bottom) antibodies indicated in **(i)** WTH and **(ii)** ΔHS-40 humanised erythroid cells (Ter119-positive cells purified by auto–magnetic-activated cell sorting). The y-axis represents enrichment over the input DNA, normalised to a control sequence in the mouse *GAPDH* gene. The x-axis represents the positions of Taqman probes used. The coding sequence is represented by the three exons (orange rectangle; Promoter/Ex1, Ex2, Ex3) of the human α-globin genes. MEx2 and MGata6 denote control sequences in the mouse genome. A deleted region in the humanised mice, 3' alpha, shows no PCR signal. Error bars represent the standard deviation from at least two independent ChIPs. This figure is adapted from Vernimmen *et al.* (2011) and De Gobbi *et al.* (2017).

1.5 Genetic editing, epigenetic editing and single cell reporters

Genetic editing has allowed researchers to generate living models with specifically designed changes in their genetic code. Recently, the development of epigenetic editing has allowed the site-specific modification of the epigenetic landscape. These techniques provide the tools to study an innumerable amount of biological questions. Single cell reporter systems can be used to monitor gene and/or protein expression in a refined manner. The combination of these techniques can allow for the isolation of cells expressing particular proteins in a variety of mutated backgrounds. In this section I will provide a description of each of these techniques.

1.5.1 Gene targeting

The first successful gene targeting experiments in ES cells and generation of mouse chimeras earned Mario Capecchi, Martin Evans and Oliver Smithies a Nobel Prize in Physiology or Medicine in 2007 (reviewed in Capecchi 2005). These experiments used recombineering techniques to modify specific genomic loci. More recently, nucleases that introduce targeted double stranded breaks (DSBs) to DNA have been developed. The earliest of these include zinc-finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN), and are targeted by protein-DNA interactions (Joung and Sander, 2013; Urnov et al., 2005). CRISPR (clustered regularly interspaced short palindromic repeat)/Cas (CRISPR-associated) is the most recent nuclease-based technology used in genetic engineering, and is instead based on RNA-DNA interactions (Sander and Joung, 2014). In this section I will describe how cellular DSB repair mechanisms are required for gene targeting to occur. I will then discuss the recombineering, TALE and CRISPR-Cas9 gene editing technologies in detail.

1.5.1.1 DSB repair

DSB repair mechanisms (Figure 1.4) have evolved in the cell to protect the genome from loss of integrity. DSBs in the genome can arise from exogenous sources, such as ionising radiation or nucleases. The collapse or stalling of a replication fork due to the presence of an inter-strand crosslink or a single stranded DNA (ssDNA) nick can also lead to DSB formation (reviewed in Chapman et al. 2012). DSBs are then repaired by one of three mechanisms in mammalian cells: (1) the error prone nonhomologous end joining (NHEJ), (2) homologous recombination (HR) or (3) microhomology-mediated end joining (MMEJ). NHEJ and HR are the two main pathways through which the genetic engineering by site-specific nucleases is mediated (Figure 1.4). NHEJ mechanisms promote the direct ligation of double stranded DNA (dsDNA) ends and occur throughout the cell cycle. This process is error prone and often results in the formation of small insertions or deletions (indels) or substitutions. HR occurs mostly during S and G2 phases of the cell cycle, when the cell is replicating its DNA and before mitosis occurs, and leads to precise homology directed repair (HDR), often between aligned sister chromatids. The DSB repair mechanisms that occur in the cell are reviewed in detail in (Chapman et al., 2012; Symington and Gautier, 2011; Wyman and Kanaar, 2006).

1.5.1.2 Recombineering and recombinase mediated cassette exchange

Traditional molecular cloning with restriction enzymes and DNA ligases uses *Escherichia coli* (*E. coli*) as a vehicle in which plasmids can be propagated. A *RecA*/*recBCD* mutant strain of *E. coli* is often used as these strains cannot quickly degrade linear DNA that enters the cell, however these strains are also recombination deficient (Yu et al., 2000). The λ bacteriophage uses homologous recombination to repair double stranded DNA breaks (Figure 1.5). A defective λ bacteriophage

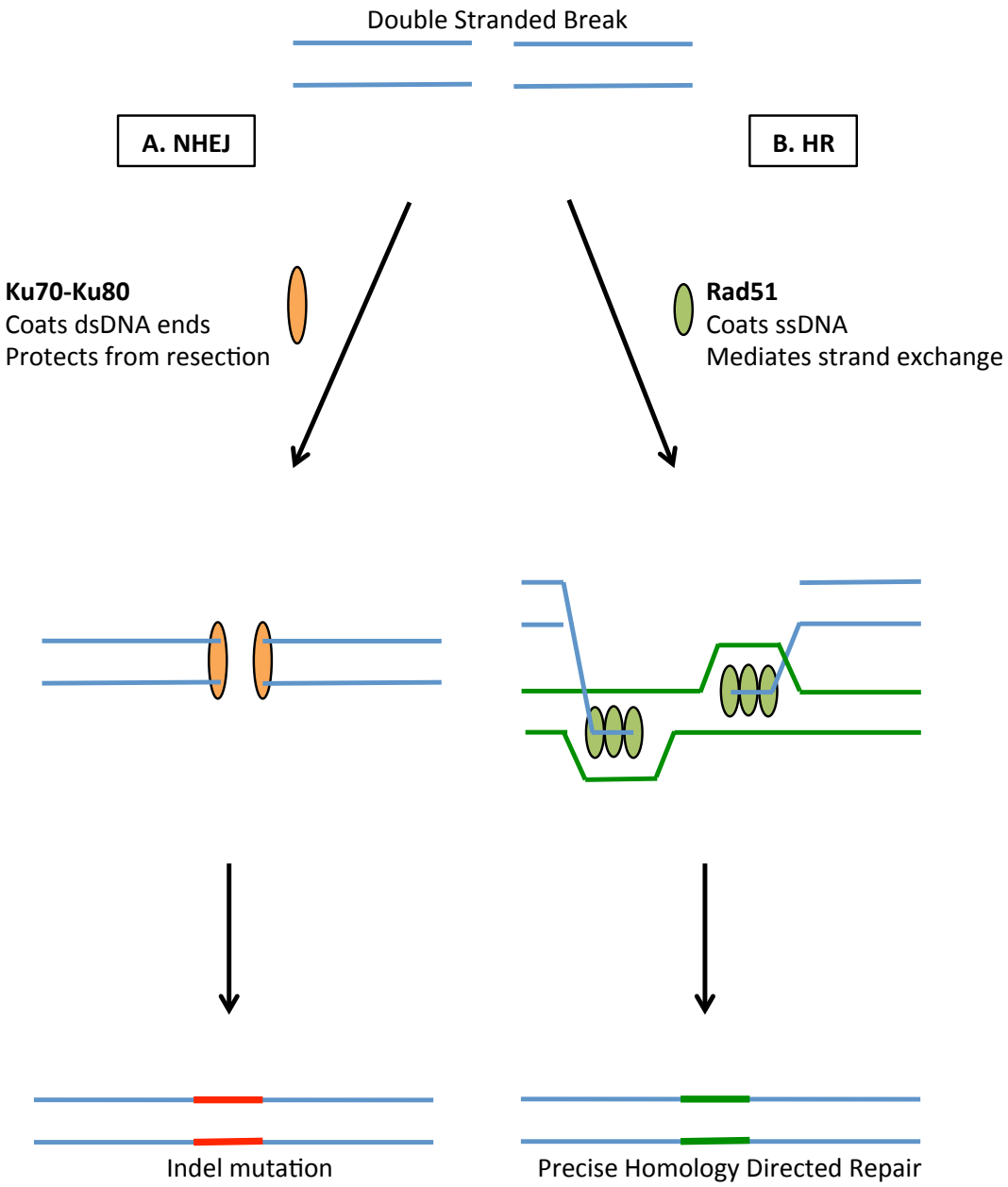


Figure 1.4 DSB repair mechanisms. Two major DSB repair mechanisms in the mammalian cell are shown along with two of the major complexes involved in these multi-protein processes.
A. The heterodimer Ku70-Ku80 coats dsDNA ends to protect them from resection and recruits the DNA ligase machinery. Error prone NHEJ then occurs, often leading to indel or substitution mutations.
B. The Rad51 protein coats ssDNA after 5` -3` resection by exonucleases and mediates strand exchange in HR. The coated 3` ssDNA strand invades a homologous template (green lines), for example a sister chromatid or a plasmid vector with homologous arms, and leads to precise HDR. This figure is adapted from Ran *et al.* (2013).

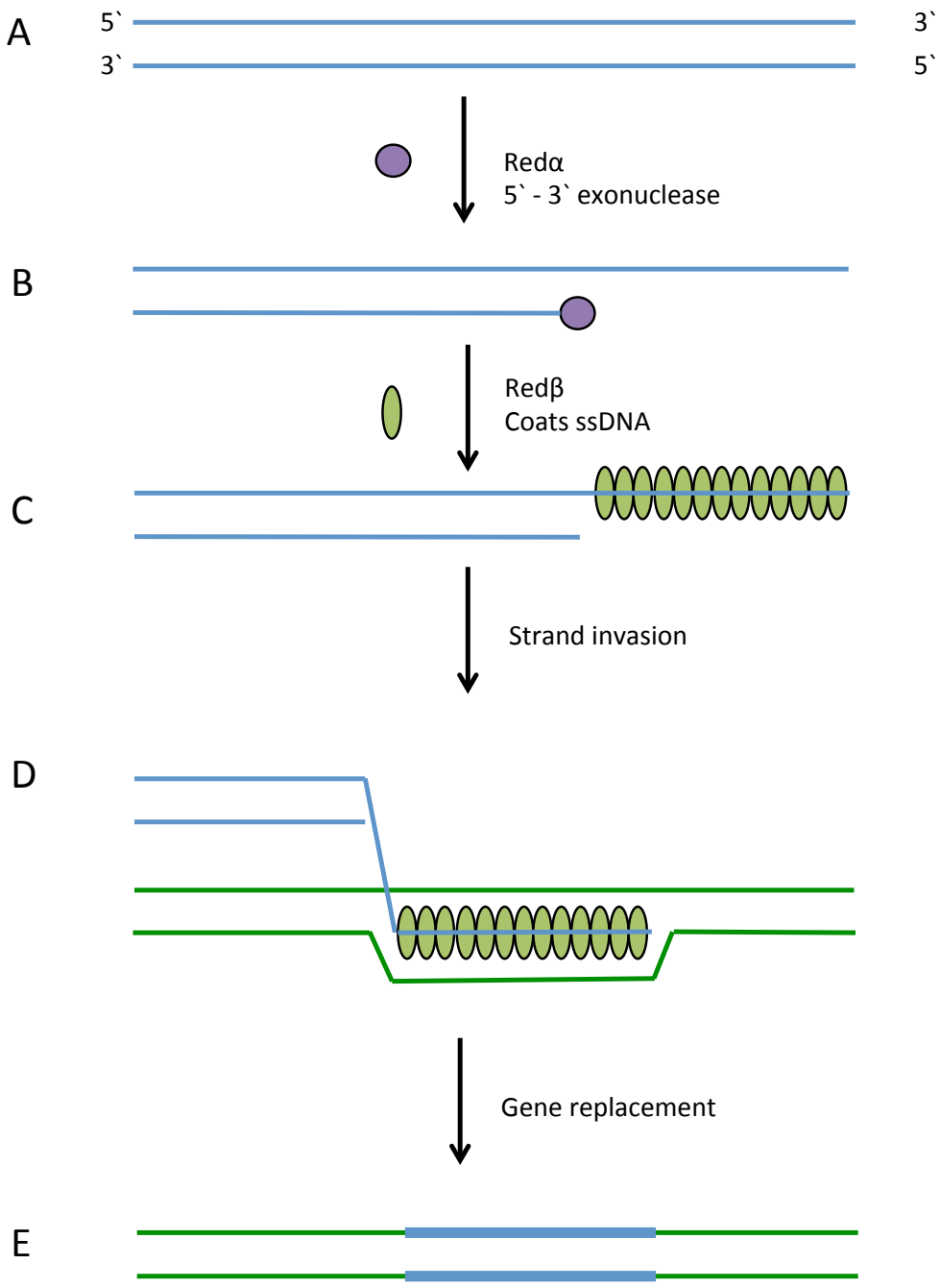


Figure 1.5 Homologous recombination using DY380 *E. coli* cells.
A. The free end of a PCR product, for example, is recognised as a DSB in the DNA.
B. Redα is a 5`-3` exonuclease from the λ bacteriophage which digests one strand of DNA at a DSB creating a single stranded overhang.
C. Redβ then binds to and coats the single stranded DNA.
D. This protein-DNA complex aligns with a homologous template (green lines), for example a BAC, and the 3` end becomes a primer for DNA replication.
E. A gene from the PCR product (blue) becomes integrated into the BAC DNA (green). This figure is adapted from Muyrers *et al.* (2000).

containing the three λ genes required for recombination, *redα*, *redβ* (also called *exo* and *bet*, respectively) and *gam*, was inserted into the DY380 *E. coli* strain that is capable of propagating BACs (Lee et al., 2001). Redα is a 5-3` exonuclease that resects the DNA at a DSB, creating a 3` single stranded DNA (ssDNA) overhang (Figure 1.5B). Redβ then coats the 3` ssDNA overhang and recombination takes place *via* strand invasion (Figure 1.5C) or single stranded annealing. The Gam protein inhibits the *E. coli* RecBCD exonuclease activity if it is present in the cell (Muyrers et al., 2000). These λ genes are under the control of the temperature sensitive λcI-repressor in DY380 cells (Lee et al., 2001). A brief activation of these genes by heat shock is sufficient to induce recombination in the DY380 cells in the presence of a homologous template (Lee et al., 2001). Strategies has been developed to manipulate sequences in BAC DNA using this recombination-mediated genetic engineering technology (recombineering) in the presence of a plasmid with short homology arms to the BAC, and a selectable marker which is inserted into the BAC (Lee et al., 2001; Muyrers et al., 2001).

The development of the Cre-*loxP*, FLP-*FRT* and Dre-*rox* recombination systems have been extremely important in developing mammalian gene targeting methods (Branda and Dymecki, 2004). The Cre, FLP and Dre site-specific recombinases (SSRs) were isolated from bacteriophage P1, *Saccharomyces cervisiae* and bacteriophage P1 related phages, respectively (Anastassiadis et al., 2009; Sauer and McDermott, 2004). Each SSR mediates recombination events between a unique pair of recombination target sites (RTSs) which are ~34 bp in length and composed of two palindromic sequences separated by ~8 bp (Anastassiadis et al., 2009). The RTSs, *loxP*, *FRT* and *rox*, can only recombine with another copy of the same RTS (Anastassiadis et al., 2009; Branda and Dymecki, 2004). Alternative RTSs have been developed in which substitutions are made within the spacer region that generate a new specific target site, compatible with the same recombinase. For example, the F_3 RTS variant of *FRT* can only recombine with other copies of F_3 but not *FRT*, and the recombination of both RTSs is mediated by FLP (Schlake and Bode, 1994). The SSRs mediate recombination between a pair of RTSs and, depending on the orientation of the RTSs, this can lead to the deletion, insertion, inversion or translocation of the intervening DNA sequence (Branda and Dymecki, 2004). This

powerful technology has been adapted for use in gene targeting strategies in which recombinase mediated cassette exchange (RMCE) allows the manipulation of chromosomal DNA in mammalian cells (Figure 1.6).

BAC recombineering is used in combination with RMCE to enable gene targeting. BAC libraries contain thousands of large genomic loci (>100 kb in length) from a variety of species. BAC recombineering can be used to insert, replace or manipulate specific genomic loci in a BAC. The resulting recombinant DNA with long, ~10 kb homology arms, can be then retrieved into homology directed repair (HDR) donor plasmids that contain RTSs for use in gene targeting strategies (Lee et al., 2013). Gene targeting plasmids contain selectable markers, RTSs, and long, ~10 kb homology arms to a genomic locus. These targeting plasmids can be transfected into mammalian cells where homologous recombination can cause the integration of the targeting cassette at a particular genomic locus (Figure 1.6A). The targeted cells can then be transfected with the HDR donor plasmid. RMCE can then occur between the compatible RTSs on the HDR donor plasmid and targeted locus in the presence of an appropriate SSR (Figure 1.6B). This leads to the replacement of the targeting cassette with the HDR cassette. Any remaining selectable markers integrated with the HDR cassette and flanked with RTSs can then be removed in the presence of an appropriate SSR (Figure 1.6C).

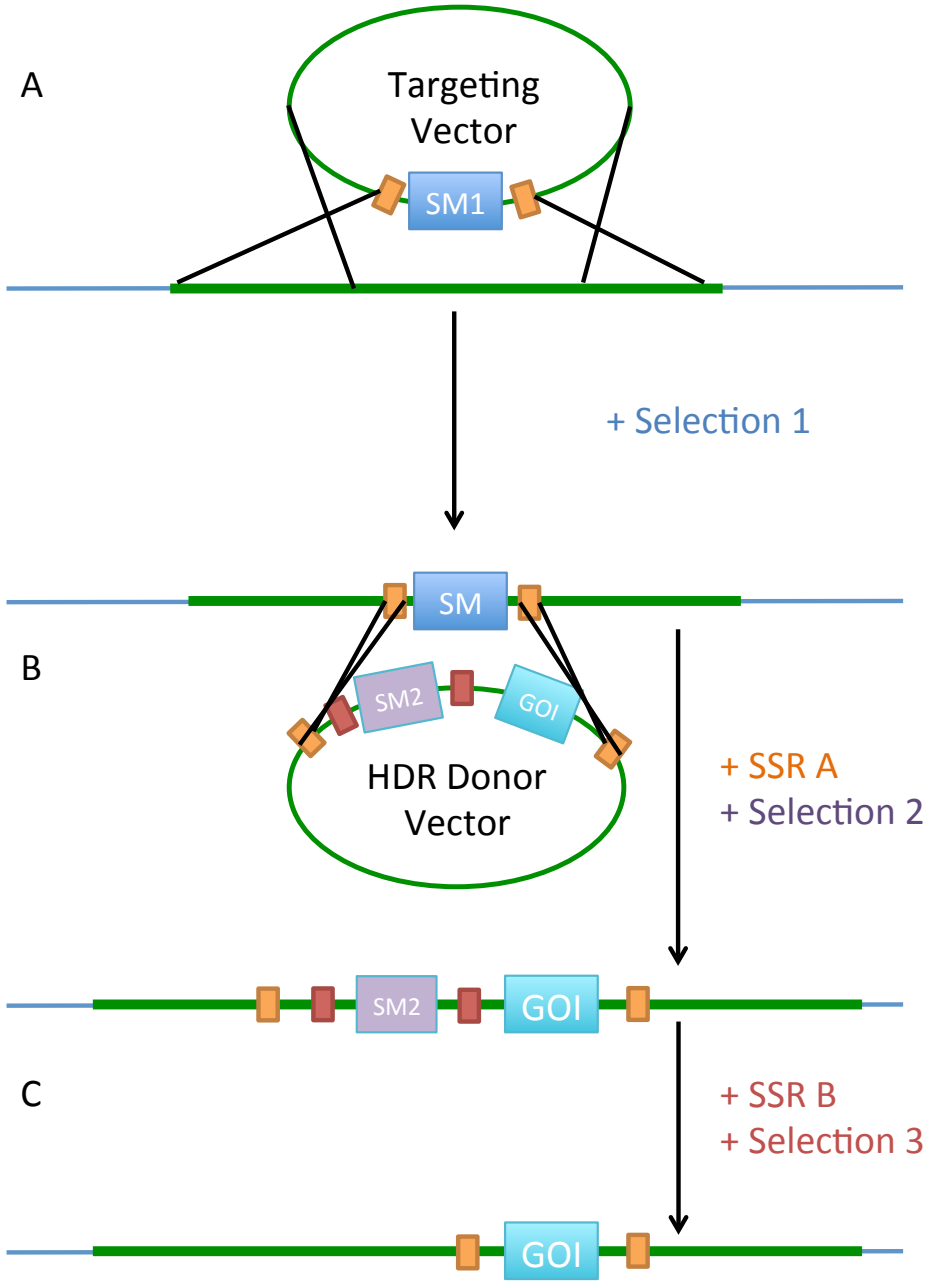


Figure 1.6 Recombinase mediated cassette exchange (RMCE) in mammalian cells.
A. Recombination occurs between the ~10 kb of homologous regions (green) in both the targeting vector (ellipse) and the genomic DNA (blue). This recombination mediates the insertion of a selection marker (SM1) flanked with RTSs (orange rectangles).
B. Recombination, in the presence of SSR A, occurs between the RTSs (orange rectangles) in the HDR donor vector (ellipse) and in the genomic DNA. This recombination mediates the exchange of the RTS flanked DNA.
C. Recombination, in the presence of SSR B, occurs between the RTSs (red rectangles) in the genomic DNA and mediates the removal of the second selection marker (SM2).

1.5.1.3 TALEN technology

TALEs were discovered in the plant pathogenic bacteria *Xanthomonas*. TALEs naturally act to trigger disease by binding specific host genes and mimicking endogenous TFs to active them. They are composed of a N-terminal domain, a central domain of tandem repeats approximately 34 amino acids in length and a C-terminal domain which includes a nuclear localisation signal and an acidic transcriptional activation domain (Boch et al., 2009). The most C-terminal tandem repeat contains only ~20 amino acids and so is termed the last half repeat (Boch et al., 2009). Within the tandem repeats, two hypervariable amino acids at position 12 and 13, termed the repeat-variable diresidue (RVD), confer specificity of target binding. Each RVD corresponds to a single nucleotide, with some degeneracy in the cipher (Boch et al., 2009; Moscou and Bogdanove, 2009). The decoding of this cipher enabled the construction of synthetic TALE arrays that could bind specific DNA target sites (Boch et al., 2009).

Many groups began fusing a synthetic TALE to the nuclease *Fok*I (TALEN), a restriction enzyme that requires dimerization to cleave DNA. Two TALENs could then be targeted to each DNA strand to generate a DSB, thus triggering the NHEJ or HR repair pathways of the cell. In this way, indels can be generated at specific locations or, in the presence of a HDR donor vector, sequences can be integrated into the target genome (Cermak et al., 2011; Christian et al., 2010; Li et al., 2011; Miller et al., 2011).

1.5.1.4 CRISPR/Cas9 technology

A recently identified RNA guided endonuclease is being used as a tool to genetically engineer almost any target sequence in a number of species with relative ease (Le Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013a). The CRISPR/Cas system is a natural prokaryotic adaptive immune system identified in a large range of bacterial and archael hosts. The CRISPRs are in part derived from foreign DNA and guide the Cas endonucleases to find and cleave invading DNA strands. Three types of CRISPR-Cas systems have been described and the most well characterized and

widely used is a Type II system derived from the *Streptococcus pyogenes* Cas protein 9 (Cas9).

Short single guide RNAs (sgRNAs) contain a scaffold and a 20 nt guide or seed sequence, which directs Cas9 to any genomic target sequence preceding a 5`-NGG or -NAG protospacer adjacent motif (PAM) (Hsu et al., 2013; Jinek et al., 2012). The PAM proximal 8-12 bp of the sgRNA appear to confer the most specificity in some studies (Hsu et al., 2013; Mali et al., 2013a) yet it seems the entire 20 bp target sites and the PAM sequence can also contribute to specificity (Pattanayak et al., 2013). The Cas9 nuclease creates a DSB in the target DNA 3-4 nt upstream of the PAM. Cas9 contains two nuclease domains; the HNH domain, which cleaves the complementary strand, and a RuvC-like domain, which cleaves the non-complementary strand (Jinek et al., 2012). The mutation of the RuvC-like domain converts the Cas9 into a nicking enzyme, Cas9n, while the mutation of both nuclease domains yields a catalytically dead enzyme, dCas9 (Le Cong et al., 2013; Mali et al., 2013a).

The CRISPR/Cas9 system is now widely used as a genome-editing tool. DSBs caused by the targeting of a single sgRNA in combination with Cas9, or by the targeting of two sgRNAs in combination with Cas9n, can be repaired *via* the NHEJ or HR pathways, leading to indel mutations or HDR (Figure 1.4). The ease and versatility of genomic engineering allows the creation of multiple knockout mutations (Hwang et al., 2013; Wang et al., 2013), targeted genomic rearrangements which model those occurring in cancer cells (Choi and Meyerson, 2014) and the site-specific insertion of DNA sequences including specific mutations, antibody tags or fluorescent genes (Ran et al., 2013a; Yang et al., 2013).

The site-specific activity of TALEN and CRISPR-Cas9 technology has provided efficient tools for genetic editing. Furthermore, many researchers have begun to use the nuclease deficient forms of this technology (TALE and CRISPR/dCas9) as DNA binding platforms to which a range of activating, repressing and chromatin modifying domains can be fused. In the following section I will discuss these novel epigenetic editing technologies and briefly describe some of the recent discoveries they have allowed.

1.5.2 Epigenetic editing

Many correlations have been made between the presence of particular epigenetic marks and changes in transcription. However, it still remains unclear whether some of these marks have a causative role in modulating expression rates or if their role is more important in maintaining a memory of a transcriptional state (see section 1.1.1, pp. 2; section 1.1.2 pp. 5). Recent technological advances have allowed the generation of site-specific alterations to the epigenetic landscape. The careful description of any resulting phenotypes could elucidate the function of these epigenetic marks. In this section, I will focus on TALE and dCas9 fusions that cause the activation or the repression of gene transcription and that, in the majority of cases, mediate the site-specific change of epigenetic marks.

1.5.2.1 Targeted transcriptional activation using DNA-binding domain fusions

A number of different activation domains have been used in DNA-binding domain fusions to mediate target gene activation (reviewed in Thakore et al. 2016). VP64 consists of four copies of the viral protein, VP16, which recruits a number of activating complexes, including histone acetlytransferases (Hall and Struhl, 2002). The p65 subunit of the NF-κB transcription factor complex and the human heat shock factor 1 (HSF1) have also been shown to mediate the transactivation of target genes (Konermann et al., 2014; Schmitz and Baeuerle, 1991). A TALE or the nuclease deficient dCas9 can be directly fused to an activation domain and recruited to a target locus mediating robust gene transactivation (Mali et al., 2013a; Miller et al., 2011). More recently, a number of technical modifications have been made that mediate much stronger gene activation with this system. The MS2 bacteriophage coat protein binds to a specific stem loop RNA structure from a viral genome (Peabody, 1993). Multiple copies of this target stem loop structure can be added to the sgRNA scaffold structure, providing a binding platform for dCas9, as well as MS2. The dCas9 and MS2 proteins can be fused to different transcriptional activators and the recruitment of these in tandem causes higher transcriptional activation than recruitment of an individual activation domain (Konermann et al., 2014). Similarly,

the “SunTag” protein scaffold can recruit multiple copies of a specifically designed antibody that can be fused to activation domains (Tanenbaum et al., 2014).

It remains unclear what the effects of site-specific targeting of these activation domains have on the underlying epigenetic modifications. The majority of studies focus on mRNA expression changes when using these activators and ignore potential changes to the epigenetic landscape. Targeting the VP64 transactivation domain to enhancers has been shown to result in some enrichment of H3K27ac and H3K4me1 at the *Oct4* distal enhancer and the β-globin HS2 enhancer (Gao et al. 2013; Hilton et al. 2015). VP64 targeting to the *Oct4* distal enhancer was also shown to enhance enrichment of H3K27ac and H3K4me3, and a depletion of H3K27me3 at the *Oct4* TSS (Gao et al., 2013). Another study used a panel of four guide RNAs targeting a promoter in cells expressing a dCas9 fused on both the C- and N-termini to VP64 (Black et al., 2016). Gene expression was induced by 100,000 fold but H3K4me3 was only enriched by 5-20 fold (Black et al., 2016). However, a mild ~20 fold induction of gene expression by a ZF–VP64 fusion targeting the gene promoter was not sufficient to induce H3K4me3 enrichment at the promoter (Cano-Rodriguez et al., 2016). From such a small number of studies it is hard to draw concrete conclusions although it seems that epigenetic changes at a gene promoter tend to occur when a transcriptional activator domain induces substantial changes to gene expression.

Other studies have taken the different approach of directly fusing epigenetic modifiers to DNA binding domains and searching for a resulting change in epigenetic landscape and gene transcription. A fusion of the catalytic core of the histone acetyltransferase p300 to dCas9 mediates the targeted enrichment of H3K27ac at sgRNA binding sites (Hilton et al., 2015). Targeted enrichment of H3K27ac at enhancers and promoters is sufficient to induce high levels of up to ~10,000 fold induction of target gene expression (Hilton et al., 2015). DNA demethylation by TET1 fusions targeted to gene promoters can also cause induction of expression at the target gene (Konermann et al., 2013; Maeder et al., 2013a). Site-specific targeting of the meiotic histone methyl transferase, PRMD9, can cause significant enrichment of H3K4me3 (Cano-Rodriguez et al., 2016). This H3K4me3

enrichment is sometimes, but not always, sufficient to cause induction of target gene expression (Cano-Rodriguez et al., 2016). This evidence suggests that H3K4me3 may only be instructive for gene transcription in a context dependent manner. In the following section I will discuss DNA binding domain fusions to repressive domains, and to epigenetic editing enzymes associated with gene repression.

1.5.2.2 Targeted transcriptional repression using DNA-binding domain fusions

The gene specific repression of expression is an important technique that can help to elucidate the specific function of a gene or be used for clinical applications. For some time, various groups have used tailored small RNAs to knock-down specific genes. More recently, DNA binding domain fusions to repression domains have been used for the same purpose (Boettcher and McManus, 2015). The simple binding of dCas9 to a gene coding region can cause an up to 1,000 fold reduction of gene expression *via* the steric inhibition of transcriptional machinery (Qi et al., 2013). The Krüppel-associated box (KRAB) domain is often used in DNA-binding domain fusions as it triggers the recruitment of a cascade of proteins associated with transcriptional repression, beginning with the KRAB associated protein 1 (KAP1). KAP1 then sequentially recruits enzymes associated with the deposition of H3K9me2/3 (SETDB1, G9A), the demethylation of H3K4 (LSD1), histone deacetylation (the NuRD complex) and DNA methylation (DNMT3A/B) (Groner et al., 2010). Fusions of the KRAB domain to ZFs, TALEs and dCas9 mediate the robust silencing of targeted genes (Gao et al., 2013; Gilbert et al., 2014; Groner et al., 2010; Klann et al., 2017). However, transient expression of these fusion proteins does not usually result in stable repression of transcription (Amabile et al., 2016). Some evidence has shown that expression of DNMT3A/B fusion proteins may support prolonged gene repression (Rivenbark et al., 2012; Stepper et al., 2016). A recent report achieved the stable repression of a target for over 40 days in a cell line following transient transfection of a combination of dCas9-KRAB, -DNMT3A and -DNMT3L fusions (Amabile et al., 2016). This repressor domain combination mediated an enrichment of H3K9me3 and CpG methylation, and depletion of H3K4me3 and PolII occupancy at the target gene (Amabile et al., 2016).

The site-specific targeting of epigenetic enzymes associated with gene repression has revealed mechanistic insights into the function of these marks. The targeted removal of H3K4me1/2 at enhancers by LSD1 fusions to DNA binding domains is sufficient to cause a reduction in the enhancer target gene expression, indicating the importance of this mark in enhancer mediated gene activation (Kearns et al., 2015; Mendenhall et al., 2013). The site-specific targeting of a number of PRC1 components results in the deposition of H2Aub and the subsequent recruitment of PRC2 components, supporting the role of PRC1 in PRC2 recruitment (Blackledge et al. 2014). A recent report tested the capability of EZH2 fusions to dCas9 in repressing gene transcription (O'Geen et al., 2017). EZH2 fusions were sufficient to cause the site-specific deposition of H3K27me3 (O'Geen et al., 2017). However, some repression was observed from fusions of catalytically inactive forms of EZH2 that could not mediate H3K27me3 deposition (O'Geen et al., 2017). Furthermore, the rate of gene repression was not amplified in the presence of H3K27me3 enrichment, and the relative amount of repression was gene target and cell type dependent (O'Geen et al., 2017). This suggests that H3K27me3 may not be instructive of gene repression, and that EZH2 may mediate gene repression by other mechanisms independent of the H3K27 methylase activity.

The use of DNA binding domains as platforms for testing the effects of catalytic domain fusions on gene expression is promising. Likewise, DNA binding platforms have been modified for use in monitoring transcriptional activity. In the following section, I will discuss these technologies and others that have been developed for visualising gene expression on a single cell basis.

1.5.3 Single cell reporter systems

Heterogeneous gene expression is regularly reported in apparently homogenous cell populations. This heterogeneity can occur within a cell; monoallelic gene expression has been reported to occur at >4,000 genes in humans (Savova et al., 2016). There is also much evidence of heterogeneous gene expression and protein accumulation between cells in populations of pluripotent and multipotent stem cells

(reviewed in Torres-Padilla & Chambers 2014; Nimmo et al. 2015). It has been suggested that this heterogeneous gene expression is a developmental mechanism required for the priming of different lineages, leading to increased differentiation diversity (Torres-Padilla and Chambers, 2014). The first description of heterogeneous gene expression arose from single cell multiplexed PCR and today has culminated in single cell RNA-seq (reviewed in Nimmo et al. 2015). Although informative, these techniques do not allow for the visualisation of heterogeneous gene expression within the cell. In this section, I will briefly describe some of the technologies that have been developed to facilitate the tracing of gene expression on a single cell basis.

Fluorescent proteins and dyes (fluorophores) are useful tools that emit light when excited at a particular wavelength. Using the appropriate combination of lasers, multiple fluorophores can be detected in tandem, allowing for multiplexing (Kremers et al., 2011). Many groups have used transgenic techniques to generate fusions of a gene of interest to a fluorescent reporter gene. For example, in frame fusions of two different fluorescent reporter genes to the start codon of β- and γ-globin were generated in the K562 erythroleukaemia cell line, which shows heterogeneous and low expression of β-globin (Voit et al., 2014). Firefly luciferase is sometimes used as a fluorescent reporter in the presence of its substrate, luciferin, as it has a shorter half-life than most other fluorescent reporter proteins (Suter et al., 2011). When transgenic techniques are not feasible, gene expression can be visualised in large populations of live cells through the use of nano-flare technology (also commercially available under the name SmartFlare; Seferos et al. 2007). Nano-flares are 13 nm gold particles to which multiple oligonucleotides that are complementary to a target RNA transcript are bound. Fluorescent dyes are bound to the oligonucleotides that are quenched as long as they remain in close proximity to the core gold nanoparticle. However, when the nano-flares enter a cell *via* receptor-mediated endocytosis, the binding of target RNA facilitates the release of the fluorescent dye (Seferos et al., 2007). This permits the user to detect, and potentially isolate, live cells that are actively transcribing the target RNA.

The above techniques allow the detection of RNA or protein expression in single cells. However, the detection of where and when transcription occurs within a

cell is also a useful method. The cellular localisation of transcriptional activity can be observed through the *in situ* hybridisation of multiple fluorescent probes to a particular RNA target (RNA-FISH; Femino et al. 1998; Bartman et al. 2016). Although useful and quantitative, RNA-FISH requires the fixation of cells before *in situ* hybridisation. This means vast numbers of cells must be analysed before useful conclusions about the temporal dynamics of transcription are made. Other methods have been developed that allow the real time imaging of transcription in a living cell. As described previously (section 1.5.2.1, pp. 40), the MS2 system can be used to mediate the site-specific localisation of the MS2 binding protein to specific RNA stem-loop structures (Peabody, 1993). Transgenic techniques can be used to integrate these binding sites into genomic DNA (Bakstad et al., 2012). This allows the visualisation of transcriptional bursting in real time *via* the binding of an MS2 protein fused to a fluorescent reporter (Bertrand et al., 1998; Fukaya et al., 2016). The PP7 bacteriophage coat binding protein binds to a different RNA stem loop sequence and so can be multiplexed with the MS2 system (Fukaya et al., 2016; Larson et al., 2011).

The use of single cell reporter systems has facilitated the advancement of knowledge in a number of areas. For example, the visualisation of transcription following enhancer looping to target genes showed that enhancers are involved in the regulation of transcriptional bursting (Bartman et al., 2016; Fukaya et al., 2016). Transgenic reporter cell lines are extremely useful tools for isolating expressing and non-expressing cells within apparently homogenous cell populations (Chambers et al., 2007). This allows for the detailed comparison of sub-populations of cells that may have underlying differences leading to heterogeneous gene expression.

1.6 Summary and outline of thesis

The humanised mouse model (WTH), in which ~120 kb of the human α-globin locus has replaced one of the mouse α-globin loci, is an important tool for studying human α-globin transcriptional regulation (Wallace et al., 2007). The deletion of the major HS-40 α-globin enhancer from the humanised locus (ΔHS-40) reduces human α-globin expression to <2% of normal levels (Vernimmen et al., 2009). It is also

known that the HS-40 plays an important role in the removal of the PcG proteins from the α-globin genes during differentiation (Vernimmen et al., 2011). Analyses of the ΔHS-40 mouse model revealed that bivalency at the α-globin genes in normal hES cells is not resolved in ΔHS-40 erythroid cells (De Gobbi et al., 2011; Vernimmen et al., 2011). Furthermore, α-globin is heterogeneously expressed in ΔHS-40 erythroid cells (De Gobbi et al., 2017). Moreover, it was proposed that this epigenetically mixed population of cells underlies the heterogeneous α-globin expression pattern in ΔHS-40 erythroid cells. If this were the case, it would reveal an important new function of enhancers in the maintenance of homogenous epigenetic landscapes *via* the consistent removal of PcG proteins in all cells.

To determine the true nature of bivalency at the α-globin locus, a reporter system for monitoring α-globin expression in live WTH and ΔHS-40 erythroid cells was required. Establishment of a reporter system in humanised mES cells would allow erythroid cells to be isolated based on their α-globin expression level. Erythroid cells could be obtained from *in vitro* differentiation of the humanised mES cells. Primary erythroid cells could also be isolated from mouse lines generated from the humanised mES cells. ChIP-qPCR analyses could then be performed on expressing vs. non-expressing ΔHS-40 erythroid cells. This would determine if H3K4me3 was the only mark found at α-globin in expressing cells, and if H3K27me3 was the only mark found at α-globin in non-expressing cells.

In this thesis, I first isolated and characterised suitable WTH and ΔHS-40 humanised mES cell subclones to perform gene-targeting experiments on. I also optimised a robust *in vitro* differentiation protocol for the generation of erythroid cells (Chapter 3, pp. 79). Next, I designed gene-targeting strategies for generating a fluorescent reporter knock-in at human α-globin 2 (*HBA2*) in the humanised mES cell subclones (Chapter 4, pp. 99). Initially, recombineering strategies were designed to achieve a knock-in using RMCE techniques. However, due to technical difficulties with BAC targeting using the RMCE strategy, and due to the recent advent of the reportedly fast and efficient CRISPR/Cas9 gene editing technologies, a CRISPR/Cas9 strategy was implemented instead. To this end, I generated and validated a number of CRISPR/Cas9 tools for generating a reporter knock-in at the *HBA2* gene (Chapter 4,

pp. 99). Using these tools, I performed a number of gene targeting attempts using a WTH humanised mES cell subclone (Chapter 5, pp. 132). A number of clones were isolated that contained a successful knock-in of the reporter cassette. This marked a significant step forward in the generation of a fluorescent reporter knock-in in a ΔHS-40 background. Future work will include the completion of the remaining gene targeting steps, generation of mouse colonies from these mES cell lines and analysis of primary erythroid cells from these mice.

While the nature of bivalency at the human α-globin genes in ΔHS-40 cells is unclear, so too is the true function of bivalency in ES cells. It has been proposed that the function of bivalency is to poise developmental genes for timely activation or permanent repression (Bernstein et al., 2006). However, some evidence to the contrary cast doubt on this hypothesis (discussed in section 1.2.4, pp. 18). The site-specific manipulation of bivalent marks in ES cells and analysis of any dynamic changes in expression during differentiation of these cells would reveal information about the function of bivalency during development. Here, I designed and developed a cloning strategy that allows the fast generation of epigenetic editing tools that can bind to specific genomic loci (Chapter 6, pp. 161). A validated TALE targeting the bivalent mouse promoter of *Nrp1* had already been published (Therizols et al., 2014) and so I generated a number of expression vectors containing TALEs that bind to the *Nrp1* promoter, and are fused to enzymatic regions involved in catalysing the deposition or removal of bivalent histone marks. I tested the effects of these TALE fusions on *Nrp1* expression and histone methylation status in humanised mES cells. Notably, one of these TALE fusions, containing the H3K27me3 demethylase *Jmjd3*, was sufficient to elicit a transcriptional response at the *Nrp1* gene. These experiments have provided an important foundation for future studies, in which bivalent histone modifications would be manipulated in ES cells and the effects of this could be monitored during *in vitro* differentiation. For example, the bivalent state at human α-globin could be manipulated in WTH mES cells and the effects of this on human α-globin expression dynamics could be monitored during *in vitro* differentiation of these cells. Finally, the results of this thesis are summarised and discussed in Chapter 7, pp. 209).

Chapter 2 Materials and Methods

2.1 DNA procedures

2.1.1 Genomic DNA (gDNA) isolation

2.1.1.1 GDNA isolation from mES cells for T7 Endonuclease I Assay

Cells were harvested 48 h after transfection and gDNA was isolated. A cell pellet was resuspended in ice-cold cell lysis buffer (140 mM NaCl, 1.5 mM $MgCl_2$, 10 mM Tris pH 8.0, 0.5% NP-40) and incubated on ice for 5 min. The lysed cells were pelleted, washed with ice-cold PBS and resuspended in 1 volume of Proteinase K Buffer (100 mM Tris pH 7.5, 12.5 mM EDTA, 150 mM NaCl, 1% SDS) with 200 µg/ml fresh Proteinase K (Promega). The proteinase K digestion was incubated at 37°C overnight. The DNA was extracted using phenol:chloroform and ethanol precipitated (see section 2.1.2, pp. 49).

2.1.1.2 GDNA isolation from mES cells for PCR screening

Frozen 96 well plates containing mES cells were defrosted at room temperature. Quick Extract Buffer (50 µl; EpiCentre) was added to each well and mixed by pipetting and transferred to a 96 well PCR plate. This plate was incubated in a thermocycler under the following conditions; 65°C for 15 min, 98°C for 8 min. A small volume of the crude extract (1-2 µl) was used in PCR screening and the extracts were stored at -20 to 4°C.

2.1.1.3 GDNA isolation from mES cells for Southern blotting

Cells at ~100% confluency in a 6 well plate were washed twice with PBS. A volume of 500 µl of lysis buffer (10 mM Tris, 10 mM EDTA, 0.1% SDS) with freshly added Proteinase K (0.1 mg/ml) was added to each well. The plates were sealed tightly with parafilm and incubated at 37°C overnight.

The following day, lysates were transferred to a 1.5 ml eppendorf tube. A volume of 1 ml of DNA isolation solution (70% acetone, 5% DMF, 25% H_2O) that had been stored at -20°C was added to each tube. The tubes were vortexed briefly

until a white DNA pellet precipitated out of solution. The pellet was transferred with a pipette tip to a new 1.5 ml eppendorf containing 0.5 ml 70% EtOH and placed on a shaker for 2 h. Samples were then centrifuged for 15 min at 13,000 rpm. The supernatant was discarded and a speedvac was used to remove as much ethanol as possible. The pellet was resuspended in 50 µl TE buffer and incubated at 55°C for 2 h. In some cases, a phenol:chloroform extraction and ethanol precipitation was then performed on the samples (see section 2.1.2, pp. 49). The concentration and purity of DNA samples was determined using a Nanodrop (Thermo Scientific).

2.1.2 Phenol chloroform extraction and ethanol precipitation of DNA

A solution of phenol:chloroform (1/1 volume) pH 8.0 (Sigma) was added to 1 volume of cell lysate after Proteinase K digestion. The samples were briefly vortexed and centrifuged for 3 minutes (min) at 16,000 rcf. The aqueous top phase containing DNA was transferred to a new tube. Chloroform (1 volume; Sigma) was added to the sample to remove any remaining phenol traces. The samples were briefly vortexed again and centrifuged for 3 min at 16,000 rcf. The aqueous top phase containing DNA was transferred to a new tube. The DNA was precipitated by the addition of 0.1 volumes of 3 M NaOAC pH 5.5 (Ambion) and 2 volumes of 100% Ethanol. The samples were incubated at -20°C for at least 30 min and centrifuged at 16,000 rcf for up to 60 min at 4°C. The supernatant was removed and the DNA pellet was washed in 1 ml 70% Ethanol. The samples were centrifuged at 16,000 rcf for 30 min at 4°C. The supernatant was removed and the DNA pellets were left to air dry completely before resuspension in an appropriate volume of nuclease free water (NFW; Sigma).

During the phenol:chloroform extraction of DNA samples for Southern blotting, the vortexing steps described above were not performed. Instead, the samples were incubated on a tube rotator for at least 20 min.

2.1.3 Polymerase Chain Reaction (PCR)

Routine PCR reactions were performed with DreamTaq Green DNA

polymerase (Thermo Fisher Scientific) or an in-house generated Taq Polymerase (Engelke et al., 1990; Pluthero, 1993). A typical 25 µl PCR reaction contained 2.5 µl 10x DreamTaq Green reaction buffer or 10x Thermopol Buffer (200 mM Tris-HCl pH 8.8, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 20 mM $MgSO_4$, 1% TritonX100), 0.5 µl dNTP (10 mM), 2.5 µl each of forward and reverse primers (10 µM), 0.125 µl DreamTaq DNA Polymerase and 10-100 ng template DNA, with the final volume made up to 25 µl with ddH_2O. PCR reactions for cloning, recombineering or the generation of Southern blot probes, etc. were performed using Q5 High Fidelity DNA Polymerase (NEB). A typical 25 µl Q5 PCR reaction contained 5 µl 5x Q5 reaction buffer, 0.5 µl dNTP (10 mM), 1.25 µl each of forward and reverse primers (10 µM), 0.25 µl Q5 High Fidelity DNA Polymerase and 1-100 ng template DNA, with the final volume made up to 25 µl with ddH_2O. Certain PCR reactions were optimized by supplementation with 1 M final concentration of Betaine (Sigma). The prepared reactions were incubated in a thermocyler according to the manufacturers instructions.

2.1.4 Real Time Quantitative PCR (qPCR)

QPCR was performed on DNA templates using either Taqman Universal PCR Master Mix (Applied Biosystems) or SYBR Green PCR Master Mix (Applied Biosystems). Briefly, for each Taqman qPCR reaction (performed in technical duplicates), 12.1 µl Taqman Universal PCR Master Mix was added to approximately 10 ng cDNA, 0.2 µM forward and reverse primers, 0.1 µM Probe labelled with 5`-6-FAM and 3`-TAMRA in a final volume of 25 µl. For each SYBR Green qPCR reaction (performed in technical duplicates), 12.5 µl SYBR Green PCR Master Mix was added to approximately 10 ng cDNA or 1 µl ChIP material and 0.2 µM forward and reverse primers in a final volume of 25 µl. Reactions were assembled on ice in a MicroAmp Fast Optical 96 well reaction plate (Applied Biosystems) and plates were sealed using a MicroAmp Optical adhesive film (Applied Biosystems). Plates were centrifuged briefly and qPCRs were performed on a 7500 Real-Time PCR System (Applied Biosystems) according to manufacturers recommendations.

Previously published primer sets were used where possible (Anguita et al., 2004; De Gobbi et al., 2011) or designed using Primer 3 online software (Untergasser et al., 2012). Designed primer sets were analysed using the IDT online OligoAnalyser Tool (PrimerQuest® program, IDT, Coralville, USA. Retrieved 12 December, 2012. http://www.idtdna.com/Scitools) and where possible followed the following restrictions; 20 bp in length, GC content 20-80%, T_m=58-60°C, no more than two G/C at 3` end, Hairpin T_m <40°C, self/hetero-dimer ΔG>-8 kcal/mole. Primer sets were tested by performing a standard curve (with technical triplicate wells) on relevant target template and confirmed to have an r-squared value of ≥0.99 and efficiency of 80-120%. PCR products were run on a 3% agarose gel to ensure a single band of the expected size was produced. A table of the primer sets used in this thesis can be found below in Table 2.1. For mouse *Nrp1* expression analysis, a PrimeTime Std qPCR Assay Mm.PT.58.9810806 for use with Taqman chemistry spanning Nrp1 exon 12-13 was used.

Table 2.1 qPCR Primer and Probe sequences.

Primer Name	Chemistry	Sequence 5` - 3`
Human UBC	SYBR	F: ATTTGGGTCGCGGTTCTTG
		R: TGCCTTGACATTCTCGATGGT
Human alpha globin (ex3)	SYBR	F: ACCTCCCCGCCGAGTTC
		R: AGGCTCCAGCTTAACGGTATTTG
HS-40	Taqman	F: CAGGCTCCAGGCCCATATC
		R: CCTCCTGCACTGTCCTTTGAC
		P: TGCCCAAGAGCTCCTTCTGCAACC
Human 5' alpha Probe	Taqman	F: GACCTCCTGGTGCTTCTGCTT
		R: CCAGCATGGATTCCAGGACTT
		P: AAACGTCCCTGTCCCCGGTGCT
Human 3' alpha Probe	Taqman	F: TCCAGGAAGCCCTCAGACTAAC
		R: CTCTAACCATCACACAAGTACACACAGA
		P: CTGGTCACCTTGAATGCCTCGTCCA
mCherry	Taqman	F: AAGACCACCTACAAGGCCAAGAA
		R: GGAGGTGATGTCCAACTTGATGT
		P: CTGCCCGGCGCCTACAACGTC

The table is continued on the next page.

Table 2.1 *Table continued.*

Primer Name	Chemistry	Sequence 5` - 3`
Human alpha globin Ex3	Taqman	F: ACCTCCCCGCCGAGTTC
		R: AGGCTCCAGCTTAACGGTATTTG
		P: TTCACCCCCGCGGTGCATG
Neomycin	Taqman	F: TGCCGAGAAAGTATCCATCATG
		R: GTTTCGCTTGGTGGTCGAAT
		P: CATACGCTTGATCCGGCTACCTGCC
Mouse UBC	SYBR	F: AGGTCAAACAGGAAGACAGACGTA
		R: TCACACCCAAGAACAAGCACA
Mouse alpha globin Ex3	SYBR	P: AGGCATCAGGGTGTCCACTTT
		F: CACAGAGGCAAGGAATTTATCCA
Mouse beta globin Ex1	SYBR	R: CCAATCTGCTCACACAGGATAGAG
		P: AGCAAATGTGAGGAGCAACTGA
Mouse alpha globin Ex2	Taqman	F: CACCAAGACCTACTTCCCTCACTT
		R: AGAGCATCGGCGACCTTCT
		P: CCACGGCTCTGCCCAGGTCAAG
Mouse CpG beta actin	Taqman	F: CGGTGTGGGCATTTGATGA
		R: CGTCTGGTTCCCAATACTGTGTAC
		P: AAGATGGACCTAATACGGCTTTTAACACCCG
Mouse GAPDH	Taqman	F: CAAGGCTGTGGGCAAGGT
		R: TCACCACCTTCTTGATGTCATCA
		P: ACGGGAAGCTCACTGGCATGGC
Nrp1 Upstream A	SYBR	F: ACAACACCTGCTCCTCTGTT
		R: TGGGAAGAAAGGGGAAGCAA
Nrp1 Upstream B	SYBR	F: ACGAGCAGCCTAGTTCAGTT
		R: GTCACTAGGGAGCAGGTTGT
Nrp1ex2A	SYBR	F: GGGTCACACGGAGGGTTTAT
		R: AGATCTCTTCCCTGCAACCA
Mouse CpG Actin	SYBR	F: CGGTGTGGGCATTTGATGA
		R: CGTCTGGTTCCCAATACTGTGTAC
Mouse Gata6	SYBR	F: GGTTGGACGTGAGGTTTTGC
		R: GAAAGTGAGATGCCAGGAAAGG
Mouse 5' alpha1	SYBR	F: ACTAAGACCAGCTGTGCCAGATG
		R: TTCCAAAGGGTGGCCAATT
Mouse 3' alpha2	SYBR	F: TGGCTCCGTGCATTTTCC
		R: CCCTTCCAGAGAGTGGGTAAGG
Mouse GAPDH	SYBR	F: CAAGGCTGTGGGCAAGGT
		R: TCACCACCTTCTTGATGTCATCA

2.1.4.1 Statistical analysis of qPCR data

The absolute Ct values from the qPCR runs were obtained from the 7500 Real-Time PCR System (Applied Biosystems). Statistical analyses of the Ct values were performed using the 7500 Real-Time PCR System (Applied Biosystems), Excel (Microsoft) or GraphPad (Prism) software packages. Briefly, the average value of the technical replicates was determined. The Ct value of a gene of interest was normalised to a reference (housekeeping) gene (ΔCt). The ΔCt values were normalised to the average of the ΔCt values of all the replicates of a reference sample (ΔΔCt). The fold change in gene expression was calculated using the following formula;

$$\text{Fold change in gene expression} = 2^{-\Delta\Delta Ct}$$

(Schmittgen and Livak, 2008). A one-way analysis of variance (ANOVA) was performed using the GraphPad (Prism) software package. A paired two-tailed t-test was performed using the Excel (Microsoft) or GraphPad (Prism) software packages. A P value of ≤ 0.05 was determined to indicate statistical significance. Dot plots were generated using the GraphPad (Prism) software package and bar charts were generated using the Excel (Microsoft) software package.

2.1.5 DNA digestion with restriction enzymes

Restriction digests were performed with recombinant enzymes obtained from either Roche, NEB or Promega using the supplied reaction buffers. DNA fragments were analysed by agarose gel electrophoresis on 0.8-3% agarose gels with TAE (Tris-acetate EDTA) buffer, stained with 1x SYBR Safe (Thermo Fisher Scientific) or 1x GelRed (Biotium) and visualised under a blue light transilluminator or UV lamp.

2.2 RNA procedures

2.2.1 RNA extraction

Where possible, a cell pellet of 5 x 10^6 cells was washed two-three times in

PBS before either snap freezing on dry ice or homogenizing in 350 µl lysis Buffer RLT (Qiagen) and then snap freezing on dry ice. Samples were stored at -80°C for later processing. Frozen samples were defrosted quickly by incubating briefly in a 37°C water bath and RNA was extracted with either an RNA Easy Mini (for samples containing >5 x 10^6 cells) or an RNA Easy Micro Kit (for samples containing ≤5 x 10^6 cells) (Qiagen) following the manufacturers recommendations. Immediately following RNA elution from the spin columns the samples were treated with DNaseI using a DNA free DNaseI kit (Invitrogen) to remove any traces of contaminating genomic or plasmid DNA. Directly following DNaseI treatment RNA samples were aliquoted to avoid multiple freeze/thaw cycles and snap frozen on dry ice. An aliquot of each sample was immediately tested on a Bioanalyzer (Agilent) using either an RNA ScreenTape or a High Sensitivity RNA ScreenTape (Agilent) to determine RNA concentration and quality. Only samples with a RNA integrity number of ≥7 were used in further experiments.

2.2.2 Reverse transcriptase

Reverse transcriptase (RT) reactions were performed using the SuperScript III Reverse Transcription Kit (Invitrogen) on RNA samples to generate cDNA templates according to manufacturers recommendations. Briefly, a reverse transcription reaction contained 2.5 µM Random Hexamers, 0.5 µg Oligo dT (12-18 nt), 5mM DTT, 1x First Strand Buffer, 1 µl RNase OUT, 1 µl SuperScript Reverse Transcriptase and equal quantities of RNA for each sample, made to a total volume of 20 µl with NFW. RT negative reactions were also performed in which the SuperScript Reverse Transcriptase was replaced with NFW.

2.3 Chromatin immunoprecipitation procedures

2.3.1 Sonication of MES cells for ChIP

A known number of mES cells (less than 8 x 10^6 cells) were pelleted in a 1.5 ml nuclease free eppendorf tube and washed two to three times in PBS. The cell pellet was resuspended in 1 ml fix solution to crosslink DNA to protein complexes (1

ml room temperature complete mES cell media supplemented with 13.16 µl Formaldehyde solution 36.5-38% (Sigma)) and incubated for exactly 15 min at room temperature on a rotating wheel. Glycine was added to a final concentration of 125 mM to quench the formaldehyde reaction and samples were incubated for 5 min at room temperature on a rotating wheel. All further steps were performed on ice. The fixed cells were pelleted by spinning at 1000 rcf for 5 min at 4°C and washed three times in ice-cold PBS. The cell pellet was then resuspended in 300 µl SDS Lysis Buffer (Millipore) supplemented with 1x complete EDTA-free Protease Inhibitor cocktail (Roche). The cell lysates were transferred to 1.5 ml Bioruptor Pico Microtubes with Caps (Diagenode) and incubated at room temperature for 2 min followed by a 10 min incubation on ice. Samples were sonicated for 7 min (30 sec on/30 sec off x 14 cycles) in a Bioruptor Pico sonication device (Diagenode) pre-cooled to 4°C. Before and after sonication, samples were warmed slightly such that the SDS had not precipitated. The sonicated chromatin samples were transferred to 1.7 ml Maxymum Recovery Snaplock Polypropylene Tubes (Axygen) and centrifuged at 16,000 rcf for 10 min at 4°C. The cleared chromatin was aliquoted into new tubes, snap frozen on dry ice and stored at -80°C.

2.3.2 Assessment of sonicated chromatin quality

A volume of 0.2 M NaCl was added to an aliquot of sonicated crosslinked chromatin. The crosslinks were reversed by heating to 65°C overnight in a thermomixer (Eppendorf) set to 1000 rpm. A phenol:chloroform extraction and ethanol precipitation of DNA was performed on samples. The concentrations of the DNA samples were determined using a nanodrop spectrophotometer and 1-3 µg of each sample was run on an agarose gel to determine sonication quality.

2.3.3 Chromatin immunoprecipitation (ChIP)

A sample of chromatin was defrosted at room temperature and diluted in ChIP dilution buffer (Millipore) supplemented with 1x “complete” EDTA-free Protease Inhibitor cocktail (Roche). Aliquots of 1 ml diluted chromatin were

transferred into 1.7 ml Maxymum Recovery Snaplock Polypropylene Tubes (Axygen). A sample of 50 µl of the diluted chromatin was stored at -20°C as an input control. A volume of 60 µl of salmon sperm DNA/Protein A/agarose slurry (Millipore) was added to the diluted chromatin and the samples were incubated at 4°C on rollers for 30-45 min in a "pre-clearing" step. The samples were centrifuged at 2000 rcf for 3 min at 4°C and the supernatant was transferred to a new pre-cooled tube. The appropriate antibody was then added (see Table 2.2 for the antibody details) to the diluted chromatin and these samples were incubated overnight at 4°C on rollers. A volume of 65 µl of salmon sperm DNA/Protein A/agarose slurry (Millipore) was added and the samples were incubated for 1 h at 4°C on rollers.

The agarose slurry was pelleted as before and washed with 1 ml for 5 min at 4°C on rollers. The slurry was washed sequentially with Low salt immune complex wash buffer (Millipore), High salt immune complex wash buffer (Millipore) and LiCl immune complex wash buffer (Millipore). The agarose slurry was then washed twice with TE (Millipore). The DNA was recovered from the agarose slurry by the addition of 250 µl freshly prepared elution buffer (0.2 µm filtered 0.1 M $NaHCO_3$, 0.4% Ultrapure SDS (National Diagnostics)) and incubation at 37°C with shaking at 1000 rpm for 15 min. The supernatant was recovered from the tube without disturbing the agarose pellet and the elution was repeated a second time.

The chromatin crosslinks of the ChIP and the input samples were reversed by the addition of 0.2 M NaCl and incubation for 4 h at 65°C. To digest the proteins, 10 mM EDTA, 40 mM Tris pH 6.5 and 20 µg proteinase K (Promega) were added to each tube and the samples were incubated for 1 hr at 45°C. A phenol:chloroform extraction and ethanol precipitation was preformed as previously described (section 2.1.2, pp. 49) with the incubation of precipitated DNA with 20 µg Glycogen (Roche) at -20°C overnight before washing of the DNA pellet. The air-dryed DNA pellet from the ChIP samples was resuspended in 40 µl NFW and the input control samples were resuspended in 200 µl NFW at 4°C overnight. An aliquot of 1 µl was used per qPCR reaction and technical duplicates were performed on all qPCR reactions.

Table 2.2 Details of antibodies used in ChIP protocols. A quantity of 2.5 µg of either antibody was added to each 1 ml sample of diluted chromatin.

Antibody	Supplier	Product Number	Lot	Species	Dilution
H3K4me3	Millipore	15-10C-E4	NG1715786	Rabbit monoclonal	1:400
H3K27me3	Millipore	07-449	DAM1641103	Rabbit polyclonal	1:400

2.4 Southern blotting procedures

2.4.1 Southern blotting

A typical 50 µl Southern blot restriction digestion reaction contained 5 µl 10x reaction buffer from supplier, 5 µl high concentration restriction enzyme (eg. 50 units/µl *Bam*HI, Invitrogen), 1 µl Spermidine (50 mM, Acros Organics) and 10 µg mES cell gDNA, with the final volume made up to 50 µl with ddH_2O. The gDNA was typically digested for 24 h with addition of 1 µl fresh restriction enzyme after approximately 8 h. The digested samples were migrated on a 0.8% agarose gel with TBE buffer (Tris-borate EDTA) precast with SYBR Safe (Thermo Fisher Scientific) at 15 V for 24-96 h and visualized using a blue light transilluminator.

The DNA was depurinated by washing the gel gently for 30 min with 0.25 M HCl. The HCl was removed and the gel was rinsed briefly with ddH_2O. The DNA was denatured by gentle shaking the gel with 1 M NaOH for 40-50 min. The NaOH was removed and the gel was rinsed briefly with ddH_2O. The gel was then soaked in neutralizing buffer (1.5 M NaCl, 0.5 M Tris-HCl pH 7.5) for 1.5 h. The neutralisation buffer was removed and the gel was rinsed briefly with ddH2O. The DNA was transferred to a positively charged nylon Amersham Hybond-XL membrane (GE Healthcare) overnight by capillary action in 6x SSC (0.9M NaCl, 90mM tri-sodium citrate dehydrate, pH 7.0). Briefly, the transfer assembly was constructed by placing an absorbent sponge in a large tray of transfer buffer. A layer of Whatman paper, the agarose gel and the nylon membrane were then placed, in that order, on top of the absorbent sponge. To draw the transfer buffer through the agarose gel, another layer of whatman paper and stacks of absorbent paper towels were placed on top of the nylon membrane. The next morning the membrane was

allowed to air dry and the DNA was crosslinked using a UV Stratalinker 2400 (Stratagene) and by baking the membrane at 65°C for 2 h.

2.4.2 Radioactive labeling of DNA probes

The probe DNA fragments were prepared by Q5 High Fidelity PCR followed by gel purification or by plasmid restriction digest followed by PCR purification. The probes were labelled with the High Prime DNA labelling kit (Sigma Aldrich). A quantity of 25 ng of double stranded probe DNA was made up to 14 µl (final volume) with ddH_2O. The probe DNA was denatured for 10 min at 99°C and snap cooled on ice for 5 min. High Prime was mixed well and 4 µl was added to the cooled probe DNA with 2 µl ^{32}P-labeled dCTP, 3000 Ci/mmol, 10 mCi/ml (Perkin-Elmer). This reaction mixture was incubated at 37°C for 30 min for labelling to occur. The labelled probe was purified from unincorporated nucleotides using an Illustra G-50 Nick Sephadex column (GE Healthcare) according to manufacturers instructions. The purified probe was denatured for 10 min at 99°C and snap cooled on ice for 5 min prior to hybridisation.

2.4.3 Southern hybridization

The hybridisation steps were performed in cylindrical hybridisation bottles using rotating hybridisation ovens. The nylon membranes were wetted in 2x SSC and prehybridised/blocked for at least 4 h while rotating at 65°C in 20 ml pre-warmed hybridisation buffer (500 mM Sodium Phosphate, 7% SDS) and 2 mg Salmon Sperm DNA (Sigma Aldrich), which was first denatured for 10 min at 99°C and snap cooled on ice for 5 min. The entire volume of the eluted purified labeled probe and 1 mg Salmon Sperm DNA were prepared for hybridisation by denaturation at 99°C for 10 min and snap cooling on ice for 5 min. Following prehybridisation, the buffer was removed and 20 ml hybridisation buffer was added to the hybridisation bottles followed by the prepared labelled probe and the Salmon Sperm DNA. The hybridisation was performed overnight while rotating at 65°C.

The following day the hybridization buffer was removed and the membranes were washed while rotating at 65°C with a range of prewarmed 0.1% SDS buffer

containing 6x to 0.1x SSC. The washed membranes were blotted with absorbent tissue and sealed tightly with saran wrap to ensure they would not dry out.

2.4.4 Autoradiography

The membranes were either exposed to a storage phosphor cassette for at least 4 h and visualised on a Typhoon phosphoimager (GE Healthcare) or exposed to conventional radiographic film (Kodak) and developed using Kodak processors (Xograph).

2.4.5 Southern membrane stripping

When required, Southern membranes were stripped by washing approximately ten times with boiling 0.1% SDS, 0.1x SSC for 3 min. The stripped membranes were blotted with absorbent tissue and sealed tightly with saran wrap to ensure they would not dry out. The stripped membranes were then exposed to a storage phosphor overnight and visualised on a Typhoon phosphoimager (GE Healthcare) the next day to ensure the labeled probe had been completely removed before reprobing that day.

2.5 Protein procedures

2.5.1 Preparation of whole cell protein lysate

Cells were harvested, counted and pellets were frozen on dry ice and transferred to a -80°C freezer for long-term storage. The frozen cell pellets were defrosted on ice and resuspended in residual PBS. To lyse the cells, a solution of 2% SDS in PBS, pre-heated to 95°C, was added to the cell suspension in a 1:1 ratio and mixed. The lysate was incubated at 95°C for 5-10 min. The sample was pipetted well and vortexed to ensure homogeneity. The protein concentration was determined using a Direct Detect Assay-free Card (Millipore) or a Micro BCA Assay Kit (Thermo Scientific) according to manufacturers instructions. Where required, the samples were diluted and stored at -80°C for later use.

2.5.2 Protein gel electrophoresis

Reducing sample buffer (0.126 M Tris pH6.8, 5% SDS, 19.7% Glycerol, 10.2% 2-Mercaptoethanol, 0.02% bromophenol blue) was added in a 1:1 ratio to whole cell protein lysates. A precast Novex NuPAGE 4-12% Bis-Tris Gel (Life Technologies) was placed in a Mini-Protean Tetra Cell Gel Electrophoresis tank (Biorad). The tank was filled with 1x MOPS SDS running buffer (Life Technologies). A quantity of 10 µg of sample in 1x Sample Buffer was loaded per well alongside 3-7 µl PageRuler Plus Prestained Protein Ladder, 10 to 250 kDa (Thermo Scientific). The gel was run for 40 min at 70 V and 80 min at 120 V.

2.5.3 Western Blotting

Following completion of protein separation by gel electrophoresis, the proteins were transferred from the gel to Immobilon-PVDF Transfer membrane (Millipore) in 1x Blot Buffer (25 mM Tris, 0.2 M Glycine, 20% Methanol) using the Mini Trans-Blot Electrophoretic Transfer Cell System (Biorad). The samples were generally transferred at a constant current of 200 mA for 2 h. The protein membrane was generally blocked overnight by incubation in 5% Milk (Oxoid) in freshly prepared TBST (0.15 M NaCl, 0.01 M Tris-HCl pH 8.0, 0.1% Tween20) with gentle shaking at room temperature.

2.5.4 Chemiluminescence

The blocked membranes were washed three times for 5 min in TBST with gentle shaking. The membrane was then transferred to a 50 ml Falcon tube and incubated for 3 h at room temperature on rollers with 5 ml diluted primary antibody (see Table 2.3 for details) in 10% synthetic milk (KPL) in TBST. The membrane was then washed three times for 5 min in TBST on rollers. The membrane was incubated for 30 min at room temperature on rollers with an appropriate diluted (1/2000) secondary antibody conjugated to horseradish peroxidase (HRP) (Dako) in 10% synthetic milk (KPL) in TBST. The membrane was then washed three times for 5 min in TBST on rollers. The HRP enzyme activity was detected by incubation of the

membrane with Pierce ECL Western Blotting Substrate (Thermo Scientific) followed by exposure in a dark room to radiographic films (Kodak) for 10 sec - 5 min. After this, the membrane was washed three times for 5 min in TBST on rollers to remove the substrate. Each membrane was then incubated for 30 min with the diluted (1/50,000) β-Actin primary antibody conjugated to HRP (Sigma) as a loading control. The HRP enzyme activity was detected as before.

Table 2.3 Details of antibodies used in Western blotting protocols.

Antibody	Supplier	Product Number	Lot	Species	Dilution
FLAG	Sigma	F7425	086M4803V	Rabbit polyclonal	1:1000
HBA	Santa Cruz	sc-31110	D2710	Goat polyclonal	1:1000
mCherry	Abcam	ab183628	GR157435-31	Rabbit polyclonal	1:2000

2.6 Molecular cloning techniques

2.6.1 Culture and storage of *E. coli*

Typically, chemically competent DH5α (prepared in house or Library Efficiency DH5α Competent *E. coli* (Thermo Fisher Scientific)) or were expanded in LB medium in liquid or agar cultures at 37°C. DY380 *E. coli* strains were also used for recombineering procedures and were expanded in LB medium in liquid or agar cultures at 30-32°C. Antibiotics, IPTG and X-gal was supplemented, as required, at the concentrations listed below in Table 2.4. Liquid cultures were initiated from a single colony of streaked bacterial agar culture. Liquid cultures were expanded in a volume of 3 ml for minipreps and maxi prep starter cultures and 200-250 ml for maxiprep cultures. Maxiprep starter cultures were incubated for 4-6 h with vigorous shaking (200-220 rpm) and, subsequently, 1 ml of starter culture was used to inoculate a full volume of maxi prep culture. Mini- and maxiprep cultures were incubated overnight with vigorous shaking (200-220 rpm). For long term storage, fresh bacterial pellets were resuspended in 15% glycerol in LB. Aliquots were divided between 1.5 ml eppendorfs, snap frozen on dry ice and stored at -80°C.

Table 2.4 Supplements used in the culture of *E. coli*.

	Concentration	
Supplement	**Stock (mg/ml)**	**Working (µg/ml)**
Chloramphenicol	34	12.5
Gentamycin	50	2 4
Ampicilin	100	100
Tetracyclin	10	Liquid: 4 Solid: 5
Spectinomycin	50	50
Kanamycin	50	50
X-gal	50	20
IPTG	100 mM	100 µM

2.6.2 Transformation of chemically competent bacterial cells

A quantity of 1-10 ng double stranded plasmid DNA or 50% of ligation mixtures were transformed into 20-50 µl DH5α *E. coli* cells. The cells were defrosted from -80°C on ice. The DNA was added and gently mixed with the cells and incubated on ice for 15-30 min in a 1.5 ml eppendorf. The transformation of DNA into cells was achieved by applying a heat shock of 42°C for 45 sec to the cells in a water bath followed by incubation on ice for 2 min. Heat shocked cells were mixed with 800 µl SOB medium and incubated for up to 1 h with vigourous shaking. A volume of 100 µl of expanded cells were spread onto LB agar supplemented with the appropriate antibiotic and incubated overnight at 37°C.

2.6.3 Miniprep BAC purification

Individual DY380 bacterial colonies were picked using a sterile pipette tip and inoculated into a 1 ml LB liquid starter culture, supplemented with the appropriate antibiotics. The starter cultures were incubated at 30°C for 4-6 h with vigorous shaking (200-220 rpm). A 10ml LB liquid culture, supplemented with the appropriate antibiotics, was inoculated with 100 µl of starter culture and incubated at 30°C overnight with vigorous shaking (200-220 rpm).

The next day, bacterial cultures were pelleted at 5,000 rpm for 5 min. The pellets were resuspended in 250 µl Buffer P1 (Qiagen) in four 1.5 ml eppendorf tubes. Next, 250 µl Lysis Buffer P2 (Qiagen) was added and the samples were mixed by inversion four to eight times. Less than 5 min after the addition of lysis buffer, 250 µl of neutralization buffer N3 (Qiagen) was added and samples were mixed by inversion four to eight times. The supernatant was cleared by two rounds of centrifugation at 13,200 rpm for 5 min and transfer of supernatant to new eppendorf tubes after each round. The DNA was precipitated by mixing the cleared supernatant with 750 µl isopropanol followed by incubation on ice for 10 min. The precipitated DNA was pelleted by centrifugation at 13,200 rpm for 10 min, washed with 70% Ethanol, pelleted again and resuspended in 20 µl TE (10 µM Tris pH 8.0, 1 µM EDTA) following airdrying of the pellet for up to 1 h.

2.6.4 Miniprep plasmid purification

Plasmid DNA was isolated from 3 ml cultures using a GeneJET Plasmid Miniprep Kit (Thermo Scientific) according to manufacturers instructions.

2.6.5 Maxiprep plasmid purification

Plasmid DNA was isolated from 200-250 ml cultures using an EndoFree Plasmid Maxiprep Kit (Qiagen) according to manufacturers instructions.

2.6.6 Colony PCR

Typically, 96 bacteria colonies were picked from an LB agar plate using individual sterile pipette tips. Each tip was touched onto the surface of a fresh LB agar master plate, supplemented with the appropriate antibiotic, and also dipped into a single tube of a 96 well PCR plate containing 50 µl ddH_2O. The PCR plate was sealed and the cells were lysed by incubation at 95°C for 10 min in a thermocycler. A volume of 2 µl of cell lysate was used as a template for screening PCRs. The master LB agar plate was incubated overnight at 30°C or 37°C and the plasmid DNA

was isolated from PCR positive clones by miniprep plasmid purification from the master plate colonies.

2.6.7 Recombineering

For recombineering experiments, an *E. Coli* strain (DY380) modified to express the recombination proteins Redα, Redβ and Gam from the λ phage in the bacterial genome was used. Briefly, a PCR product was generated from a plasmid containing the desired antibiotic selection cassette using primers that introduced two 50 bp homology arms to the DNA sequence of the BAC where the insertion is required. The purified PCR product containing the insertion cassette was digested with *Dpn*I for 1-4 h at 37°C to degrade any methylated original plasmid template. The digested PCR products were run on an agarose gel and the desired fragments were gel purified. A volume of 30 µl of fresh liquid culture of DY380 cells containing a BAC was added to 1.4 ml LB in 1.5 ml ependorf tubes with pierced lids. The cultures were incubated at 32°C with vigourous shaking (1000 rpm) on a Thermomixer C (Eppendorf) until the optical density at 600 nm reached ~0.6 (approximately 105 min). The recombinase genes were induced by immediate incubation at 42°C for 15 min. The cells were then transferred to the cold room on ice where all the following steps were carried out using materials and equipment which had all been pre-cooled to 4°C.

The cells were transferred to new 1.5 ml eppendorfs and pelleted at 13,000 rpm for 30 sec. The supernatant was discarded and cells were resuspended in 1 ml ddH_2O. The cells were pelleted, resuspended and pelleted again as before. The final cell pellet was resuspended in 20 µl ddH_2O and 5 µl of 100 ng/µl *Dpn*I digested and the purified PCR product was added and mixed gently with cells. A volume of 25 µl of the cells/DNA mixture was transferred to electroporation cuvettes with 0.1 cm electrode gap (BioRad). The electroporation was performed using a Micropulsar with bacterial settings (BioRad). Immediately following electroporation, the cells were collected by washing gently with 1 ml SOC medium (Invitrogen) and incubated at 30°C with vigorous shaking (200 rpm) for 1 h. A volume of 100 µl of

electroporated cells was then spread onto LB agar plates, supplemented with the appropriate antibiotic, and incubated at 30°C for 24-48 h.

2.6.8 Purification of DNA fragments

DNA fragments were purified following agarose gel electrophoresis by visualization of the correct band using a blue light transilluminator and excision of the desired fragment using a sterile razor blade. The DNA was purified from the gel fragment using a GeneJET Gel Extraction Kit (Thermo Scientific) according to manufacturers instructions. DNA amplicons or restriction fragments were typically purified from any contaminants using a GeneJET PCR Purification Kit (Thermo Scientific) according to manufacturers instructions. Briefly, both of these techniques use silica membranes within spin columns to capture the DNA allowing washing with 70% ethanol before elution in ddH_2O or TE.

2.6.9 Ligation of DNA fragments

A linearised plasmid vector DNA was dephosphorylated with Calf Intestinal alkaline Phosphatase (CIP) or recombinant Shrimp Alkaline Phosphatase (rSAP; NEB) at 37°C for 1 h to prevent plasmid religation. Double stranded insert DNA was phosphorylated using T4 Polynucleotide Kinase (T4 PNK; NEB) in the presence of 1 mM ATP for 30 min at 37°C. The vector and insert DNA fragments were purified using a PCR cleanup kit (see above). The DNA concentration was determined by visual comparison to a standard on an agarose gel. Subsequently, 50-100 ng vector DNA was ligated at 16°C overnight with a 3-20x molar excess of insert DNA using 2 µl 10 mM ATP, 2 µl 10x T4 DNA Ligase Buffer, 2 µl T4 DNA Ligase (NEB) in a total volume of 20 µl made up with ddH_2O. The next day, 10 µl of the ligation reaction was transformed into chemically competent *E.coli* cells.

2.6.10 CRISPR sg-RNA Cloning

The CRISPR sgRNA oligos were designed using an online tool,

crispr.mit.edu. The single stranded DNA oligos containing the sgRNA sequence were synthesised by Eurogentec and cloned into Cas9 plasmids as previously described (Ran et al., 2013a). Briefly, sgRNA oligos were phosphorylated with T4 Polynucleotide Kinase (NEB), annealed in a thermocycler and digested with *Bbs*I (NEB). The double stranded digested oligos were then ligated into the plasmid pSpCas9(BB)-2A-GFP-(PX458), received as a gift from Dr. P. Hohenstein (University of Edinburgh), which was also digested with *Bbs*I. The ligation reactions were transformed into chemically competent *E.coli* and the isolated plasmids were screened by Sanger sequencing.

2.6.11 T7 endonuclease I assay

WTH3 mES cells were transfected by nucleofection with (i) a construct containing Cas9 and sgRNA-A and (ii) a vehicle control containing transfection reagent and TE. Positively transfected cells from (i) were sorted by FACS 24 h after nucleofection. The cells were harvested 48 h after transfection and gDNA was isolated from the cells. A primer set was designed to amplify a DNA region of approximately 500 bp that was predicted to undergo mutation via CRISPR/Cas9 activity. A PCR reaction was performed using Phusion High Fidelity Polymerase (NEB). The PCR reaction was purified and eluted in 50 µl EB buffer (Qiagen). A volume of 18 µl purified PCR product was mixed with 2 µl Buffer 2 (NEB) and reannealed using the following cycling parameters in a thermocycler; 95°C 7 min, ramp to 85°C at 2°C/sec, 85°C for 30 sec, 12 cycles of (85°C for 30 sec (decrease 5°C/cycle), ramp to 80°C at 0.3°C/second, 80°C for 30 sec (decrease 5°C/cycle)). Next, 10 µl reannealed PCR product was immediately digested with 0.5 µl T7 Endonuclease I (NEB), 0.5 µl Buffer 2 (NEB) and 4 µl ddH_2O at 37°C for 15 min. An aliquot of 10 µl of undigested purified PCR product was run on a 3% agarose gel stained with 1x GelRed (Biotium) alongside 15 µl of T7 digested purified PCR product. The gels were imaged and the relative band intensities were estimated using Image J software. The indel frequency was estimated as previously described using the formula

$$100 \times (1-(1-(b+c)/(a+b+c))^{1/2})$$

where “a” is the integrated intensity of the undigested PCR product and “b” and “c” are the integrated intensities of each digested product (Ran et al., 2013b).

2.6.12 Gibson assembly

A 10µl aliquot of Gibson Assembly master mix (100 mM Tris-HCl, 10 mM $MgCl_2$, 0.2 mM each dNTPs, 0.5 U Q5 High Fidelity DNA Polymerase (NEB), 0.16 U 5’ T5 exonuclease (Epicentre)) was thawed on ice. The DNA plasmid template (50-100 ng) was added along with equimolar amounts of the other assembly DNA components and the volume was made up to 20 µl with ddH_20. The Gibson assembly reaction was incubated at 50°C for 1 h and then stored on ice. A volume of 3 µl of the reaction mix was transformed into chemically competent DH5α cells as described above (Gibson et al., 2009).

2.6.13 Site directed mutagenesis (SDM)

Site directed mutagenesis primer pairs were designed using previously outlined recommendations (Zheng et al., 2004). Briefly, the targeted mutations were included in both the forward and reverse primers and were located up to 4 bp away from the 5` primer terminus and at least 6-8 bp away from the 3` terminus. At least 8 non-overlapping bases were included at the 3` end of each primer with at least one G or C nucleotide at each terminus. A mutation content of up to 17.5% was included in each of the primers.

A typical 50 µl SDM Q5 PCR reaction contained 10 µl 5x Q5 reaction buffer, 1 µl dNTPs (Promega; 10 mM), 1 µl each of the forward and reverse primers (10 mM), 0.5 µl of Q5 High Fidelity DNA Polymerase, 100 ng template DNA with the final volume made up to 50 µl with ddH_2O. A PCR reaction with 18 amplification cycles was performed in a thermocycler. The PCR reactions were purified and digested with the restriction enzyme *Dpn*I for 1-2 h at 37°C to degrade any methylated original plasmid template. *Dpn*I was deactivated by incubation at 80°C for 20 min. The digested PCR products were run on an agarose gel to confirm

successful PCR amplification and 1 µl of the digested PCR product was transformed into DH5α cells.

2.6.14 Golden gate assembly of TAL plasmids

The TALEs were designed using the online tool TAL Effector Nucleotide-Targeter 2.0 (Doyle et al., 2012). The TALE array plasmids were constructed using previously described protocols (Cermak et al., 2011). Briefly, a 20 µl reaction was prepared with 150 ng each of plasmids containing RVD 1-10 and RVD 11-20 mixed with 1 µl of the Type IIS restriction enzyme *Bsa*I (10 U, NEB), 1 µl T4 Ligase (2000U, NEB), 1x T4 Ligase Buffer (NEB) and either 150 ng pFusA or pFusB, respectively. This reaction was incubated in a thermocycler for 1-10 cycles of 37°C for 5 min and 16°C for 10 min followed by 50°C for 5 min and 80°C for 5 min. Next, 1 µl of 10 mM ATP and 1 µl of Plasmid Safe nuclease (Epicentre) were then added and the reaction was incubated in a thermocycler for 37°C for 1 h. A volume of 10 µl of this reaction mixture was transformed into DH5α cells and the transformed cells were plated on LB agar plates prepared with 0.1 mM IPTG, 20 µg/ml X-gal and 50 µg/µl Spectinomycin. The plates were incubated at 37°C overnight (and sometimes for another 24 h at 4°C to intensify the X-gal staining) and approximately 10 white colonies were picked and screened by restriction digestion and Sanger sequencing.

A modified version of pTAL-2 (Cermak et al., 2011), hereafter referred to as the TALE-Fusion-Expression vector (pTFE) was generated by molecular cloning techniques. Following screening, the TALE array plasmids, the appropriate Last Repeat Variable Domain plasmid (Cermak et al., 2011), pTFE and the kanamycin resistant subcloning plasmids were digested overnight with an excess of Fast Digest *BsmB*I (Thermo Scientific) in 1x T4 Ligase Buffer (NEB). The complete digestion was confirmed by visualising an aliquot of the digested product on an agarose gel. Next, 25 ng of digested pTFE and equimolar amounts of all the other plasmids containing the inserts required in the final construct were mixed with 1 µl of T4 Ligase (2000 U, NEB) and 1x T4 Ligase Buffer (NEB) in a final volume of 20 µl. The ligation reaction was incubated at 16°C overnight and 10 µl of this reaction was

transformed into DH5α cells. The transformed cells were plated on LB agar plates prepared with 0.1 mM IPTG, 20 ng/ml X-gal and 100 μg/μl ampicilin. The plates were incubated at 37°C overnight (and sometimes for another 24 h at 4°C to intensify X-gal staining) and white colonies were picked and screened by PCR, restriction digestion and Sanger sequencing.

2.7 Mammalian cell culture

2.7.1 MES cell culture conditions

Unless otherwise described, mES cells were cultured in Glasgow MEM (BHK-21; Life Technologies) supplemented with 10% batch tested Foetal Calf Serum (FCS; Globalfarm; batch tested for culture with mES cells), 2 mM L-Glutamine (Life Technologies), 0.1 mM MEM non-essential aminoacids (Life Technologies), 1 mM Sodium Pyruvate (Life Technologies), 0.1 mM 2-Mercaptoethanol (Life Technologies) and typically 0.2% Leukemia Inhibitory Factor (LIF) prepared by Shahida Sheraz (University of Edinburgh). At times, complete mES cell culture media was supplemented with 100 U/ml penicillin-streptomycin (Life Technologies). Tissue culture treated plates (Corning) were prepared for mES cell culture by coating with sterile 0.1% gelatin (Sigma) in ddH_20 and incubation at 37°C for at least 15 min. Gelatin was aspirated off before the addition of cells/culture media.

The mES cells were typically split 1:5 every 48 h and incubated at 37°C, 5% CO_2 in humid conditions. Briefly, cell monolayers were washed twice with sterile 1x PBS (Oxoid) and then incubated for 3 min at 37°C with TrypLE Express (Life Technologies). The enzymatic reaction was quenched by the addition of cell culture media containing 10% FCS. The cells were harvested and centrifuged at 1000 rcf for 3 min. The cell pellet was resuspended with 1 ml fresh media with pipetting at high velocity to make a single cell suspension. The cells were diluted to the desired concentration and transferred to gelatinised plates (Corning).

For long-term storage, the fresh mES cell pellets were resuspended in 10% DMSO (Sigma) in FCS and aliquoted into Nunc cryotubes (Thermo Scientific).

These were placed in polystyrene boxes in a -80°C freezer for at least 24 h. The frozen cells were transferred to a -150°C.

2.7.2 Plat-E cell culture conditions

Platinum-E (Plat-E) cells, a modified Human Embryonic Kidney 293T (HEK293T) cell line (Morita et al., 2000), were received as a gift from Dr. P Hohenstein (University of Edinburgh). Plat-E cells were cultured in Dulbecco's MEM (Life Technologies) supplemented with 10% FCS (Life Technologies), 2 mM L-Glutamine (Life Technologies), 10 μg/ml Blasticidin (InvivoGen) and 1 μg/ml puromycin (Sigma).

Plat-E cells were typically split 1:10 every 72 h and incubated at 37°C, 5% CO_2 in humid conditions. Briefly, cell monolayers were washed twice with sterile PBS and then incubated for 3 min at 37°C with Trypsin (Life Technologies). The enzymatic reaction was quenched by the addition of cell culture media containing 10% FCS. The cells were harvested and centrifuged at 1000 rcf for 5 min. The cell pellet was resuspended with 1 ml of fresh media with pipetting at high velocity to make a single cell suspension. The cells were diluted to the desired concentration and transferred to new tissue culture plates.

For long-term storage, fresh Plat-E pellets were resuspended in 10% DMSO (Sigma) in FCS and aliquoted into Nunc cryotubes (Thermo Scientific). These were placed in polystyrene boxes in a -80°C freezer. These cells were transferred to a -150°C after at least 24 h.

2.7.3 Counting cells using a haemocytometer

Cell numbers were estimated by counting with a glass haemocytometer and coverslip. Briefly, the cells were mixed to ensure a homogenous distribution and then serially diluted where required to an appropriate factor such that ~10-100 cells were visible when loaded in each set of 16 corner squares on the haemocytometer. A volume of 10 μl of diluted cells were mixed 1:1 with Trypan Blue Solution, 0.4%

(Gibco) and 10 µl of this mix was loaded on the haemocytometer. The cells were viewed on a microscope under a 10x objective lens. The live, unstained cells were counted in each of the four sets of 16 corner squares on the haemocytometer grid. The estimated number of viable cells/ml was calculated using the following formula;

(average viable cell count in 16 corner squares) x 10,000.

2.7.4 Mammalian cell transfection procedures

2.7.4.1 MES cell nucleofection

The mES cells were split 24 h before nucleofection. On the day of nucleofection, Nucleofector solution (NS) was freshly prepared by adding P2 Supplement (Lonza) to the Nucleofector Solution (Lonza) in a 1:4.5 ratio. A quantity of 3 µg DNA was added to NS to a final volume of 30 µl. The mES cells were harvested and counted and 2 x 10^6 cells were washed once in PBS and the resulting pellet was resuspended in 70 µl NS. This cell suspension was added to the DNA/NS solution and mixed with pipetting. A volume of 100 µl was transferred to an amaxa certified cuvette (Lonza). The cuvette was closed and the cells were transfected using the programme A-023 on a Nucleofector (Lonza) machine, which is optimized for use with E14 cells. A volume of 500 µl of prewarmed media was immediately added to the cuvette and the cells were plated in a gelatinised T25 flask.

2.7.4.2 MES cell electroporation

The mES cells were split 24 h before electroporation. On the day of electroporation, the mES cells were harvested and counted and 1 x 10^7 cells were washed once in PBS and the resulting pellet was resuspended in 400 µl PBS. A quantity of 100 µg of plasmid DNA at a concentration of 10 µg/µl was added to the cells and mixed with pipetting and 500 µl of the mixture was transferred to a pre-cooled electroporation cuvette with a 0.4 cm electrode gap (Biorad). The cuvette was closed and cells were electroporated (3 µF, 0.8 kV) using an electroporator optimized for E14 cells (Biorad). The electroporated cells were incubated for 15 min on ice and

were plated in a gelatinised T75 flask.

2.7.4.3 MES cell lipofection

The mES cells were split 24 h before lipofection. On the day of lipofection, 4 µg sterile plasmid DNA was diluted with Optimem (Life Technologies) to a total of 200µl. In a separate tube, 9 µl Lipofectamine 2000 (Life Technologies) was mixed with 191 µl Optimem (Life Technologies). Both of these tubes were vortexed and incubated at room temperature for 10 min. The contents from both tubes were mixed thoroughly together, vortexed and incubated at room temperature for at least 30 min. Meanwhile, 8 x 10^5 cells were washed three times in PBS and plated in 2 ml complete mES cell media without penicillin-streptomycin in a gelatinised well of a 6 well cluster plate. A volume of 400 µl of the DNA:Lipofectamine mixture was added to this well and the contents were mixed with gentle shaking. The cells were incubated for 4-6 h and the media was replaced with fresh complete mES cell media containing penicillin-streptomycin where required.

2.7.4.4 Plat-E cell transfection

Plat-E cells were split 24 h before transfection and plated at a density of 5 x 10^5 cells per well in a 6 well plate in 3 ml complete medium. A quantity of 3.3 µg of plasmid DNA was diluted with Optimem (Life Technologies) to a total of 155 µl. A volume of 9.9 µl of FuGENE HD reagent (Promega Corporation) was added directly to the diluted plasmid DNA and mixed by vortexing briefly. The solution was incubated at room temperature for 10 min. Meanwhile, the Plat-E cells were washed gently with sterile PBS and 3 ml of fresh complete media was added to each well. The 150 µl of DNA/FuGENE HD complex was gently added to each well. The tissue culture plates were rocked gently before placing in the incubator overnight. The media was replaced with fresh media 24 h after transfection.

2.7.5 MES cell targeting: picking clones

The WTH3 mES cells were transfected with equal quantities of a HDR

targeting vector and a vector containing sgRNA-A and Cas9 using one of the methods described above. Two-three days following transfection, the cells were harvested and typically 1 x 10^6 cells were plated per gelatinised 10 cm tissue culture treated dish in complete mES cell media supplemented with 1 μg/ml puromycin dihydrochloride (Sigma). The selection was maintained until colonies were picked seven to ten days later.

The media was aspirated from the dish and cells were washed twice with PBS. A volume of 10 ml PBS was added to the dish to hydrate the cells during picking. Individual mES cell colonies were picked with a sterile fine tip Pasteur pipette and transferred into a single well of a 96 "V" well plate (Costar) containing 30 μl of TrypL-E (Life Technologies). Once a full plate was picked, the cells were incubated at 37°C for 5 min. A volume of 70 μl complete mES cell media was added to each well using a multi-channel pipette and the total amount was transferred to a gelatinised 96 well tissue culture plate containing 100 μl media per well. The next day, the media was aspirated off and a fresh 200 μl of media was added. When the majority of wells in a plate were 70-80% confluent, the plate was split 1:1 into two new 96 well plates. Once these new plates were 70-80% confluent, one was used for PCR screening and the other was either split again or frozen down in 10% DMSO/FCS.

2.7.6 MES cell karyotyping

The mES cells were split such that they would be 70-80% confluent in one well of a 6 well cluster plate after 24 h. The media was aspirated off the cells and Colcemid (Life Technologies) was added 1:100 in fresh media to stop cells at the metaphase stage and the cells were incubated at 37°C for 2 h. The cells were harvested and pelleted. The cell pellet was resuspended in 5 ml 0.56% KCl (Fisher Chemical) pre-warmed to 37°C and incubated in KCl for 12 min at room temperature. A volume of 100 μl of fresh fixative (3:1 methanol: acetic acid) pre-cooled to -20°C was added dropwise with constant agitation to prevent cell clumping. The cells were pelleted by centrifugation at 1000 rcf for 3 min and

supernatant was removed. The cell pellet was resuspended in the residual buffer and 1 ml of fixative was added dropwise with constant agitation. The cells were incubated on ice for 30 min, pelleted and resuspended in 1 ml of fixative as before. An aliquot of 10-12 µl of fixed cells was dropped onto uncoated glass slides (VWR) from a height and left to dry. The metaphase spreads were stained for 15 min with 1:44 Giemsa Stain (Gibco) in water and then washed three times for 5 min in water. Once the slides were dried they were mounted with glass coverslips (Academy) using Pertex Mounting medium (Cell Path Ltd.) and left to dry overnight. The numbers of chromosomes per cell were counted from 10-50 metaphase spreads per cell clone using a high-powered 60x microscope (Leica DMRB fluorescent upright microscope with Hamamatsu digital camera).

2.8 *In-vitro* differentiation of mES cells

2.8.1 EB formation for random differentiation

The mES cells were split such that they would be 70-80% confluent after 24 h. The cells were harvested, counted and 1 x 10^7 single cells were plated (Day 0) in non-gelatinised bacterial grade sterile petri dishes (Thermo Scientific) in 10 ml of complete mES cell culture media without LIF (-LIF media) to promote spontaneous differentiation. After 48 h (Day 2) the cells were gently collected in a 50 ml falcon tube and allowed to settle in a tissue culture incubator for 10-15 min. The supernatant was removed and cells were gently resuspended in 10 ml fresh -LIF media and plated on fresh non-gelatinised bacterial grade sterile petri dishes. The media was changed again on Day 4 and 6. On Day 8, approximately twenty embryoid bodies (EBs) were carefully collected using a P1000 pipette and transferred to a gelatinised well in a 6 well cluster plate in fresh –LIF media. After two-seven days, the differentiated cells were imaged and EB beating was observed. The media was removed and the cells were washed gently with PBS. The cells were incubated with 1 ml of modified Leishman Eosin methylene blue solution (Merck) for 5 min. A volume of 1 ml ddH_20 was added and the solution was incubated for an additional 5 min, removed and left to dry overnight before imaging.

2.8.2 Suspension method

The mES cells were defrosted on Day -6 in complete mES cell media. The cells were split as usual on Day -4 in complete mES cell media. On Day -2, fresh Iscove's Modified Dulbecco's Medium (IMDM) was prepared by supplementing Glasgow MEM (BHK-21; Life Technologies) with 1.5 x 10^{-4} M Monothioglycerol (Sigma), 10% heat inactivated FCS (Globalfarm; batch tested for use with mES cells; the defrosted FCS was incubated in a 56°C water bath for 60 min with shaking every 10 min), 0.2% LIF, 100 U/ml penicillin-streptomycin (Life Technologies) and 2 mM L-Glutamine (Life Technologies). The mES cells were harvested and counted and 3 x 10^5 cells were aliquoted into one well of a 6 well cluster plate in IMDM on Day -2. Unless otherwise stated, EB-Differentiation (EB-Diff) media was prepared on Day 0 by supplementing Glasgow MEM (BHK-21; Life Technologies) with 3 x 10^{-4} M Monothioglycerol (Sigma), 10% heat inactivated FCS (Globalfarm; batch tested for use with mES cells; defrosted FCS was heat inactivated by incubation in a 56°C water bath for 60 min with shaking every 10 min unless otherwise stated), 2 mM L-Glutamine (Life Technologies), 50 µg/ml Ascorbic Acid (Sigma), 0.3 mg/ml Transferrin from human serum (Roche), 20x PHFMII (Life Technology), 100 U/ml penicillin-streptomycin (Life Technologies). On Day 0 the cells were harvested, counted and 4 x 10^4 cells in a singe cell suspension were plated on non-gelatinised 10 cm bacterial grade plates with 10 ml EB-Diff medium. The cells were incubated for 7-8 days before harvesting (see section 2.8.6, pp. 77) and imaging.

2.8.3 Hanging drop method

The mES cells at standardized passage number were defrosted on Day -6 in complete mES cell media. The cells were split as usual on Day -4 in complete mES cell media. On Day -2, fresh IMDM was prepared as above. The cells were counted and 6 x 10^5 cells were resuspended in 20 ml IMDM on Day -2. Using a multi-channel pipette, 10 µl droplets (300 cells per droplet) were pipetted onto the upturned lid of a 120 mm sterile square petri dish (Greiner). PBS (10 ml) was added to the

bottom of the petri dish so that the cells would remain hydrated. The lid was turned over and replaced back on top of the dish and placed in the incubator for 48 h.

The EB droplets were collected on Day 0 and transferred to a 50 ml falcon tube. The EBs were centrifuged at 800 rpm (80 g) for 3 min. The supernatant was discarded and the EB pellet was gently resuspended in 20 ml EB-Diff Media and transferred to an ungelatinised 10 cm bacterial grade plate. On Day 1, 20 ml EBs were gently collected from the plate and split evenly between two gelatinised T75 tissue culture flasks. Fresh EB-Diff Media (30 ml) was added to each flask. On Day 7/8 cells were imaged and harvested.

2.8.4 Supplemented EB-Differentiation media

The EB-Differentiation media was supplemented when described with 1 U/ml recombinant human erythropoietin (R&D Systems) prepared in 0.1% tissue culture grade BSA (Sigma) in PBS, 50 ng/ml recombinant human Stem Cell Factor (PeproTech) prepared in 0.1% tissue culture grade BSA (Sigma) in PBS or 1 μM Dexamethasone (Sigma) prepared in 0.1% tissue culture grade BSA (Sigma) in PBS.

2.8.5 Combined method

The mES cells at a standardized passage number were defrosted on Day -6 in complete mES cell media. The cells were split as usual on Day -4 in complete mES cell media. On Day -2, fresh IMDM was prepared as described above. The cells were counted and 3 x 10^5 cells were split into one well of a 6 well cluster plate in IMDM on Day -2. On Day 0, 6 x 10^5 cells were resuspended in 20 ml EB-Diff Media. Using a multi-channel pipette, 10 μl droplets (300 cells per droplet) were pipetted onto the upturned lid of a 120 mm sterile square petri dish (Greiner). PBS (10ml) was added to the bottom of the petri dish so that the cells would remain hydrated. The lid was turned over and replaced back on top of the dish and placed in the incubator for 48 h. EB droplets were collected on Day 2 and transferred to a 50 ml falcon tube. The EBs were centrifuged at 800 rpm (80 g) for 3 min. The supernatant was discarded and the

EB pellet was gently resuspended in EB-Diff Media and 200 EBs per plate were transferred to an ungelatinised 10 cm bacterial grade plate in 10 ml EB-Diff Media. The EBs were imaged and harvested on Day 7/8.

2.8.6 Harvesting erythroid cells from suspended EBs

On Day 7/8 of differentiation in the suspension or combined method, red blood islands were visible within large EBs. The EBs were collected in a 15 ml falcon tube and centrifuged at 1000 rcf for 3 min. The supernatant was discarded and the resulting pellet was washed twice in PBS. The supernatant was discarded and the pellet was resuspended in 3 ml Trypl-E (Life Technology). The cells were incubated for 3 min with regular agitation in a 37°C water bath. A volume of 7 ml of 10% heat inactivated FCS in PBS was added to quench the Trypl-E. The cells were passed through a 21G x1½" syringe three times to complete the EB disaggregation. The cells were counted and aliquoted appropriately for downstream purposes.

2.9 Flow cytometry

2.9.1 Fluorescent activated cell sorting (FACS)

The cells were prepared for FACS by harvesting and resuspending in 1 ml FACS Media (10% FCS in PBS). The cells were passed through a cell strainer (Falcon) to prevent clogging in the machine. The cells were kept on ice until they were sorted by Bob Fleming or Graeme Roberston (University of Edinburgh) on a FACS Aria IIIu 4-laser/11 detector Cell Sorter (BD Biosciences) and directly pooled into 15 ml tubes containing normal mES cell media. The sorted cells were then plated under normal cell culture conditions.

2.9.2 Ter119 antibody staining and flow cytometry

The cells were harvested and 1 x 10^5 cells, washed twice in PBS, were aliquoted in 1.5 ml eppendorf tubes. If a dead cell stain was being used, Zombie Yellow (BioLegend) was diluted in PBS 1:500 and 1 x 10^5 cells were resuspended in 100 µl of this working solution and incubated for 15-20 min on ice, protected from

light. A quantity of 0.5 µg Ter119-FITC Rat-anti-mouse antibody stain (BD Pharmingen) was directly added to the cells following this incubation. If a dead cell stain was not being used, 1 x 10^5 cells were resuspended in 100 µl of PBS and 1 µl Ter119 antibody stain was directly added to this cell suspension. The antibody stain was incubated for 40-60 min on ice, protected from light. The cells were washed twice in PBS and resuspended in 500 µl FACS media. The cells were kept on ice until they were analysed using a LSR Fortessa (BD Biosciences).

Chapter 3 Validation of mES Cell Lines and Optimisation of *In Vitro* Differentiation

3.1 Introduction

The isolation of mouse embryonic stem cells (mES cells) from the inner cell mass of blastocysts and development of cell culture protocols to maintain them *in vitro* was a significant advancement in the study of developmental genetics (Evans and Kaufman, 1981; Martin, 1981). In 2007, a Nobel prize for Physiology or Medicine was awarded to the principle investigator that led this research, along with researchers that completed the first site directed gene targeting of mES cells (Doetschman et al., 1987; Thomas and Capecchi, 1987). It had also been demonstrated that mES cells could be injected into developing blastocysts and transplanted into the uterine horn of pseudopregnant mice to produce chimeras that could be bred to homogeneity (Gossler et al., 1986). The combination of these findings paved the way for the generation of genome edited mouse models.

This study aimed to use gene editing techniques to knock-in a 2A-mCherry cassette in two previously engineered E14TG2a mES cell lines humanised at the α-globin locus, with a wild type sequence (Wild Type Humanised; WTH) or a deleted enhancer (ΔHS-40) (Wallace et al., 2007). The engineering of these two lines was described in Chapter 1 (pp. 28). The mES cell lines carrying the 2A-mCherry cassette within the α-globin locus would allow the monitoring of α-globin expression in erythroid cells with these two genotypes. Primary erythroid cells could be isolated from the generation of mouse lines by blastocyst injection. Erythropoiesis could also be studied *in vitro* through the *in vitro* differentiation of the mES cell lines.

A euploid karyotype is important for mES cell differentiation and mouse generation. Indeed, mES cell lines with a higher percentage of euploid cells (40 chromosomes in diploid mES cells) are better able to generate chimeras and contribute to germline transmission than lines with a high percentage of cells of abnormal karyotype (Longo et al., 1997; Suzuki et al., 1997). A study of 540 mES cell lines from 20 different research institutes in Japan showed that 33.5% of these lines had abnormal karyotypes (Sugawara et al., 2006). The proportion of aneuploid

cells also increases with the number of passages (Gaztelumendi and Nogués, 2014; Longo et al., 1997). This has significant implications when a number of different gene targeting steps are required to generate new cell lines, as the cells must be kept in culture for several passages. A trisomy of chromosome 8 is the most commonly observed karyotype abnormality, and has been shown to confer a growth advantage to cells harbouring this karyotype (Gaztelumendi and Nogués, 2014; Kim et al., 2013; Sugawara et al., 2006).

More recently, mice have also been generated from haploid mES cells (Leeb et al., 2012; Zhong et al., 2015). Genetic engineering using haploid mES cells can improve the efficiency of 'clean' targeting as only one copy of each chromosome must be targeted and there is no possibility of generating mutations in the second chromosome copy. As only one copy of the mouse α-globin locus is humanised in the cell lines used in this study, these targeting experiments enjoy the same advantage as targeting a single locus in haploid cells.

ES cells can be differentiated *in vitro* towards a number of different lineages. Once leukaemia inhibitory factor (LIF) has been removed from ES cell media, the cells can be induced to form embryoid bodies (EBs). EBs can be formed in a number of different ways including the culture of cells in suspension, in hanging drops and in semisolid media, as reviewed in Kurosawa (2007). Correctly formed EBs consist of spontaneously differentiated cells mimicking the endoderm, ectoderm and mesoderm germ layers (Kurosawa, 2007). Furthermore, EBs can be directed towards a specific developmental pathway through the use of particular media additives. *In vitro* differentiation of mES cells has long been used to study the process of erythropoiesis. The sequential appearance of primitive followed by definitive erythrocytes during *in vitro* differentiation closely mimics the erythropoiesis pathways observed in the developing embryo (Keller et al., 1993; Ma et al., 2008; Weiss and Orkin, 1996). Several reports highlight the importance of particular cytokines, including Stem Cell Factor (SCF) and Erythropoietin (Epo), and nuclear hormone receptors, including the Dexamethasone (Dex) responsive receptor, in the *in vitro* expansion of human erythroid progenitors and in the *in vitro* differentiation of mES cells towards the erythroid lineage (Carotta et al., 2004; Panzenbock et al.,

1998). EBs can also readily differentiate into other tissue types. For example, spontaneous EB contraction or "beating" and molecular characterisation indicates the formation of cardiomyocytes (Boheler et al., 2002). Also, small vascular-like channels through which red blood cells can travel can form in EBs (Wang et al., 1992).

More recently it has been shown that haematopoiesis can also be induced in human induced pluripotent stem cells (iPS cells) by the same *in vitro* differentiation methods used on human ES cells (hES cells; reviewed in Lim et al. 2013). Extensive sequential cocktails of recombinant cytokines and small molecules have also been used to produce high numbers of embryonic erythroid cells in serum and feeder-cell free systems, however the ability of these cells to achieve enucleation is severely limited (Olivier et al., 2006). The latest methods in expansion of human erythroid cells *in vitro* can produce high, scalable numbers of erythroid cells in which up to 30% of the population becomes enucleated, a marker of fully mature reticulocytes (Trakarnsanga et al., 2017). The development of such methods has important clinical implications for the possible *in vitro* production of functional red blood cells for use in transfusions (Giarratana et al., 2011; Migliaccio et al., 2012).

In this chapter, the karyotype of the WTH and ΔHS-40 mES cell lines was characterised and euploid subclones of these two lines were isolated. A robust and reproducible *in vitro* differentiation method for generating erythroid cells was optimised using the WTH3 subcloned cell line. The suitability of the humanised WTH3 subcloned cell line as a model for gene targeting and the study of human α-globin expression *in vitro* was confirmed.

3.2 Results

3.2.1 Karyotyping and subcloning of humanised mES cell lines

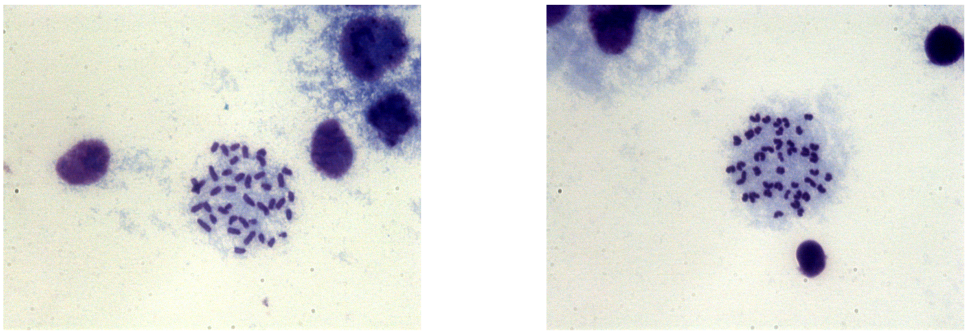
The WTH and ΔHS-40 humanised mES cells were received as a gift from Dr. Andrew Smith (University of Edinburgh). These cell lines were expanded and their karyotype was determined by counting the number of chromosomes in

approximately 25 metaphase spreads per clone (Figure 3.1A). A karyotype of 40 chromosomes for at least 70% of cells is optimal for mouse diploid ES cells to be used in gene targeting and blastocyst microinjection. The WTH clone was determined to have a karyotype of 40 chromosomes in approximately 50% of cells with approximately 20% of cells showing a karyotype of 41 chromosomes (Figure 3.1B). The ΔHS-40 clone was determined to have a karyotype of 40 chromosomes in nearly 30% of cells with nearly 50% of cells showing a karyotype of 41 chromosomes (Figure 3.1B).

To obtain cell lines with improved karyotypes (i.e. a higher proportion of cells with 40 chromosomes), the WTH and ΔHS-40 cell lines were subcloned. Briefly, cells from both lines were plated at low density (50-100 cells/plate) in 10 cm diameter tissue culture plates. The cells were left to expand for approximately 8 days and 50 ES cell colonies were picked from each line and expanded. Karyotyping assays were performed on 10 WTH subclones (WTH1-10; Figure 3.2A) and 20 ΔHS-40 subclones (ΔHS-40_1-20; Figure 3.2B). In the subclone WTH3, approximately 80% cells had a karyotype of 40 chromosomes. Nearly 60% of cells in the subclone ΔHS-40_20 showed a karyotype of 40 chromosomes with only 10% of cells in this clone showing a karyotype of 41 chromosomes.

In order to validate the ability of the mES cell subclones WTH3 and ΔHS-40_20 to randomly differentiate, an EB formation and random differentiation assay was performed on these clones. After plating of EBs, the differentiated cells were observed growing out from the larger EB mass in both cell lines (Figure 3.2C and D). The morphology of these differentiated cells is markedly different from that of undifferentiated mES cells (Figure 3.3i and ii). Beating EBs were also observed for both cell lines indicating cardiac muscle formation.

A



B

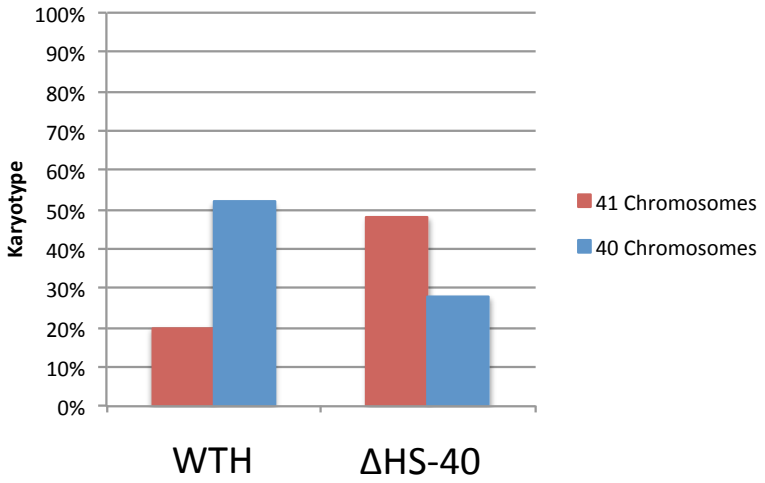


Figure 3.1 Karyotyping of WTH and ΔHS-40 mES cell lines.
A. Two examples of Giemsa stained metaphase spreads that were counted for karyotyping are shown. Images were taken using a Leica DMRB fluorescent upright microscope and Hamamatsu digital camera at 63x magnification.
B. Karyotyping results of WTH and ΔHS-40 mES cells from approximately 25 metaphase spreads.

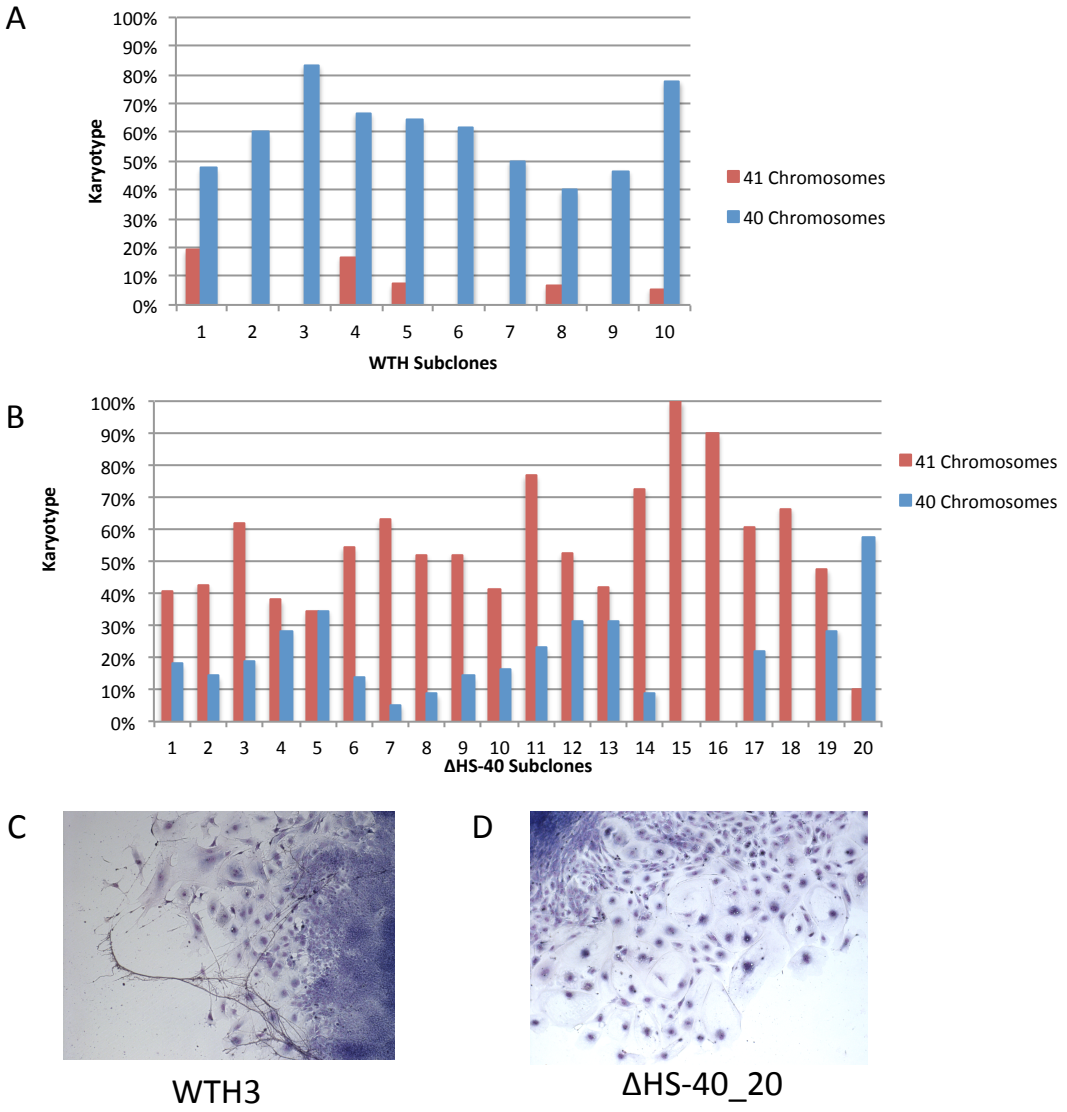


Figure 3.2 Karyotyping of WTH and ΔHS-40 mES cell subclones and EB differentiation assay.
(A-B) Karyotyping results for subclones WTH1-10 and ΔHS-40_1-20 from approximately 25 metaphase spreads per clone. Subclone WTH3 and ΔHS-40_20 were determined to show the best karyotype of >80% and >50% of cells carrying 40 chromosomes, respectively, and were used in subsequent experiments.
(C-D) Example of Leishman's staining of WTH3 and ΔHS-40_20 cells following EB formation differentiation assay show random differentiation of mES cells into a variety of cell lineages. These images were taken with a Zeiss axiovert 25s inverted microscope at 20x magnification.

3.2.2 Optimising *in vitro* differentiation protocols

3.2.2.1 Suspension method of *in vitro* differentiation

The analysis of erythroid cells derived from *in vitro* differentiation of mES cells formed an important aim of this thesis project. To this end, it was important to first develop robust and reliable *in vitro* differentiation protocols that would consistently yield high numbers of erythroid cells. The first *in vitro* differentiation protocol tested, here termed the 'suspension method', was adapted by the D. Higgs lab (WIMM, University of Oxford) from a previously published protocol (Ma et al., 2008). Briefly, mES cells were defrosted on Day –6, split at least once (Figure 3.3i), cultured for one passage in IMDM-ES media (Figure 3.3ii) and then plated in a single cell suspension in EB differentiation media in un-gelatinised plates and allowed to form EBs for approximately 7 days (Figure 3.3iii) before harvesting for downstream analysis.

The suspension method of *in vitro* differentiation was used on WTH3 cells, and various other clones, in 7 independent biological replicates (data not shown). Using this method, it was possible to recover EBs with visible red blood islands on occasion (Figure 3.3iii). However, the cell numbers recovered and number of EBs containing visible red blood islands was extremely variable between mES cell clones (Table 3.1). In most replicates, either very low cell numbers ($\leq 10^5$ cells/dish) or no cells were recovered at the harvesting step (4/7 replicates), with high variability in the appearance of red blood islands (Table 3.1).

A number of media additives are available to augment *in vitro* differentiation of hES cells. With the expectation that such additives might also augment *in vitro* differentiation of mES cells, erythropoietin (Epo) only and Epo in combination with either dexamethasone (Dex) or stem cell factor (SCF) was added to EB differentiation media and tested using the suspension method. These media supplements were tested in combination with EB differentiation media using normal foetal calf serum or serum that had been heat inactivated. Although red blood islands were recovered from all these conditions (Figure 3.4i-iv), there was no obvious difference in the intensity of the red blood islands observed in the harvested cell pellet between the different conditions (Figure 3.4C). However, the numbers of cells

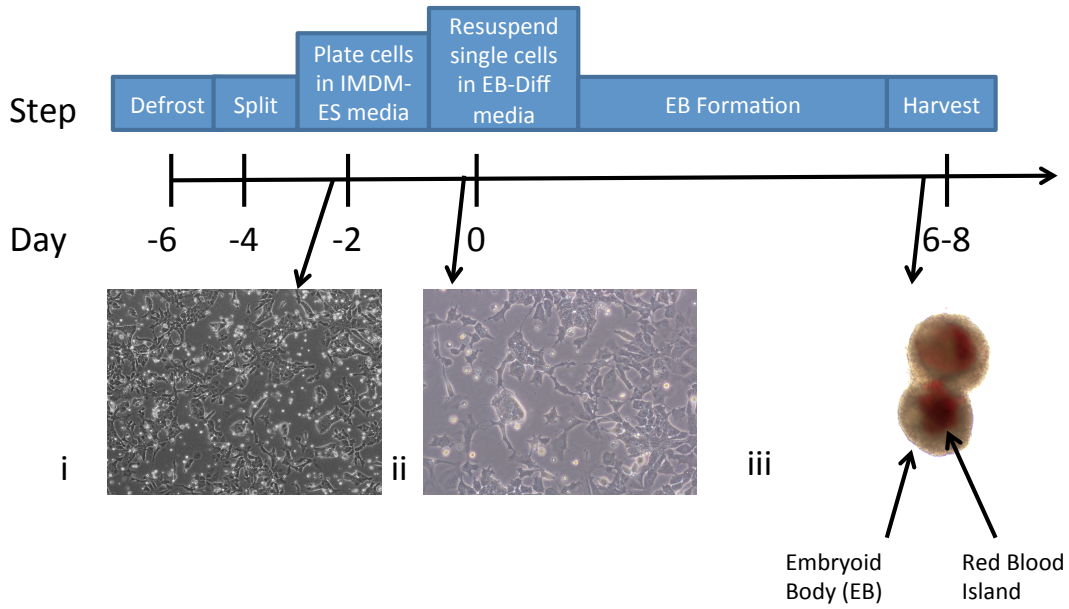


Figure 3.3 Schematic of *in vitro* differentiation using the suspension method.
(i) An example of WTH3 mES cells at day -2 before harvesting and plating in ES-IMDM media. The image was taken at 10x magnification.
(ii) An example of WTH3 mES cells at day 0 before harvesting and resuspension in EB-differentiation medium. The image was taken at 20x magnification
(iii) An example of two WTH3 EBs at day 7 containing red blood islands fused together. The image was taken at 10x magnification.
All images were taken at using a Zeiss axiovert 25s inverted microscope.

recovered at the harvesting step from treatments with heat-inactivated serum were consistently much higher than treatments with normal serum (Table 3.2).

Table 3.1 Cell numbers and blood island scoring of mES cell clones.
Results from mES cell clones including WTH3, ΔHS-40_20 and 6 other clones generated from gene targeting experiments in Chapter 5 are shown. Cell numbers harvested from one single 10 cm dish of cells differentiated *in vitro* with the suspension method are shown. A scoring legend indicating the number of EBs that contained visible red blood islands per dish is also shown.
Red Blood Island Scoring Legend
- No red blood islands observed
+/- ≤5 red blood islands observed
+ 6-10 red blood islands observed
++ >10 red blood islands observed

Clone	Passage number at Day -6	Cell count x 10^5	Blood islands
WTH3	13	1	-
ΔHS-40_20	6	60	++
A5	19	4	-
B1	19	4	+/-
EA1	9	4	+
EB7	9	9	+
EB12	9	12	++
ED9	9	17	++

Table 3.2 Cell numbers recovered from a 10 cm plate following *in vitro* differentiation of WTH3 mES cells using the suspension method with media additives. A greater number of cells were recovered at the harvesting step from *in vitro* differentiation experiments using heat-inactivated serum.

	Cell Number x 10^6	
Media Variant	**Normal Serum**	**Heat-Inactivated Serum**
Original	0.43	2.25
+Epo	0.2	2.2
+Epo +Dex	0.55	1.2
+Epo +SCF	0.53	1.9

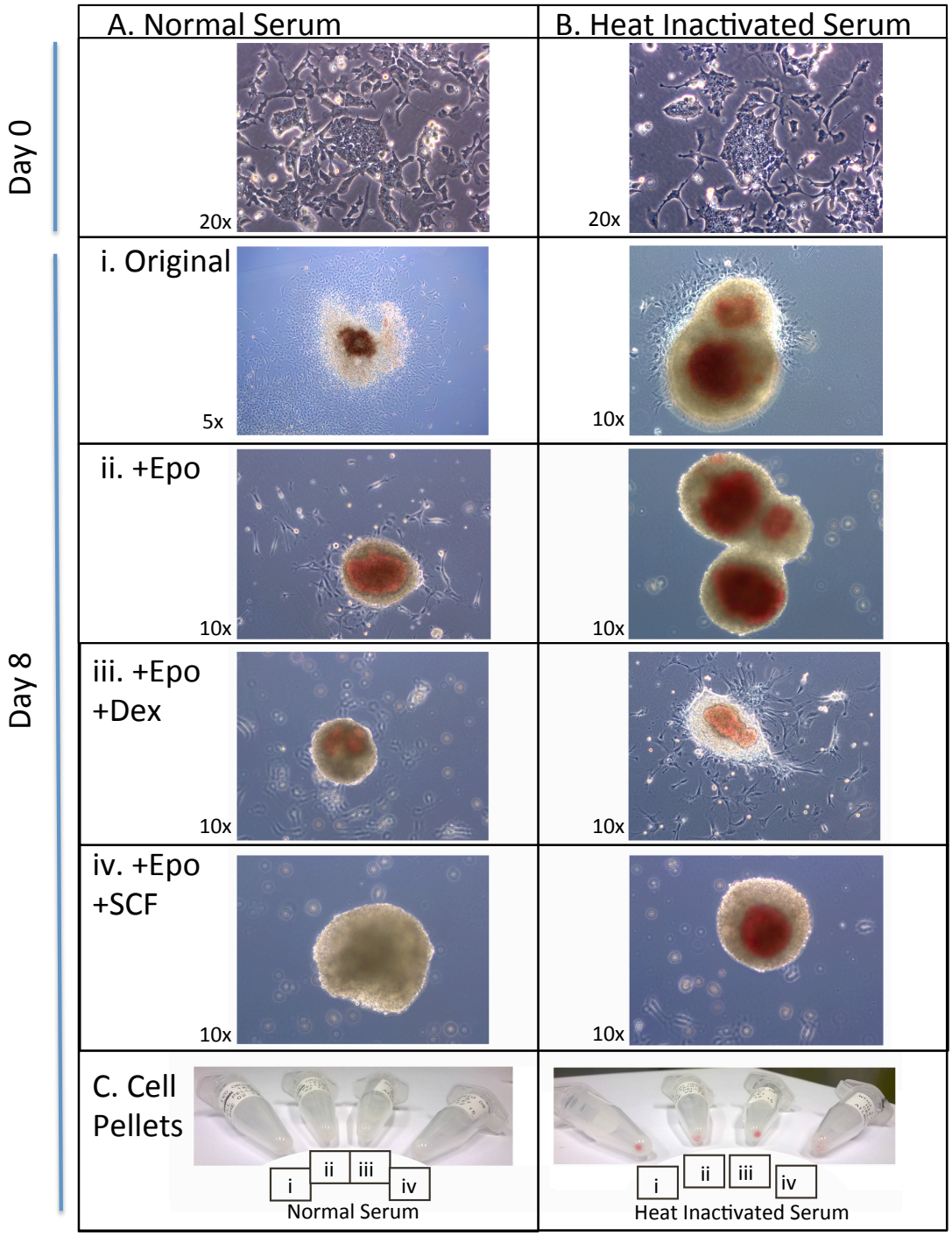


Figure 3.4 *In vitro* differentiation of WTH3 mES cells using the suspension method with media variants.
(A) Normal vs (**B)** heat inactivated serum was used in combination with **(i)** original media and media with additives **(ii)** Erythropoietin, **(iii)** Erythropoietin + Dexamethasone and **(iv)** Erythropoietin + Stem Cell Factor. These images were taken with a Zeiss axiovert 25s inverted microscope at the magnification indicated in the figure. *The figure legend is continued on the next page.*

3.2.2.2 Hanging drop method of *in vitro* differentiation

Due to a lack of reproducibility using the suspension method for *in vitro* differentiation, a new protocol was tested, termed the "hanging drop method". This protocol was developed through discussion with the L. Forrester lab (CRM, The University of Edinburgh). This method differed from the suspension method in that EBs were formed on Day -2 from a cell suspension in hanging drops of IMDM-ES media. Each EB was formed from 300 single cells in 10-20 µl media suspended in a droplet hanging from the lid of a sterile petri dish. These EBs were harvested and cultured in suspension in EB-Differentiation media for 24 hours before transfer to gelatinised tissue culture plates in which they were allowed to adhere to the plate surface before harvesting after approximately 7 days (Figure 3.5). The hanging drop method was tested using normal EB-Differentiation media and supplemented media, using both normal and heat inactivated foetal calf serum. Although the hanging drop method produced EBs of a consistent and reproducible size and shape (Figure 3.6, Day 0), none of the treatments tested were sufficient to yield any visible red blood cells (Figure 3.6i-iv).

Figure 3.4 *Figure legend continued:*
Images of eppendorf tubes containing harvested cell pellets are shown in **(C)**. The heat inactivated serum appears to give a more vibrant red cell pellet but the media additives do not appear to improve red blood island formation beyond the red cell intensity observed in the cell pellet from heat inactivated serum in combination with (i) the original method.

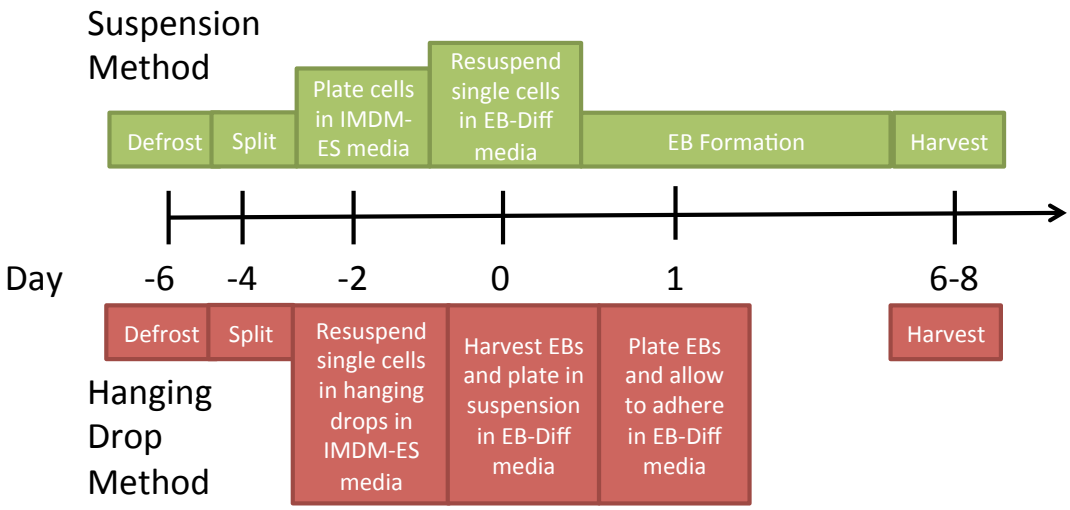


Figure 3.5 Schematic comparing *in vitro* differentiation using the suspension vs. the hanging drop method.
The suspension method (green) differs from the hanging drop method (red) in the way EBs are formed (day 0 in suspension method vs. day -2 in hanging drop method) and allowed to grow in suspension vs. adherence.

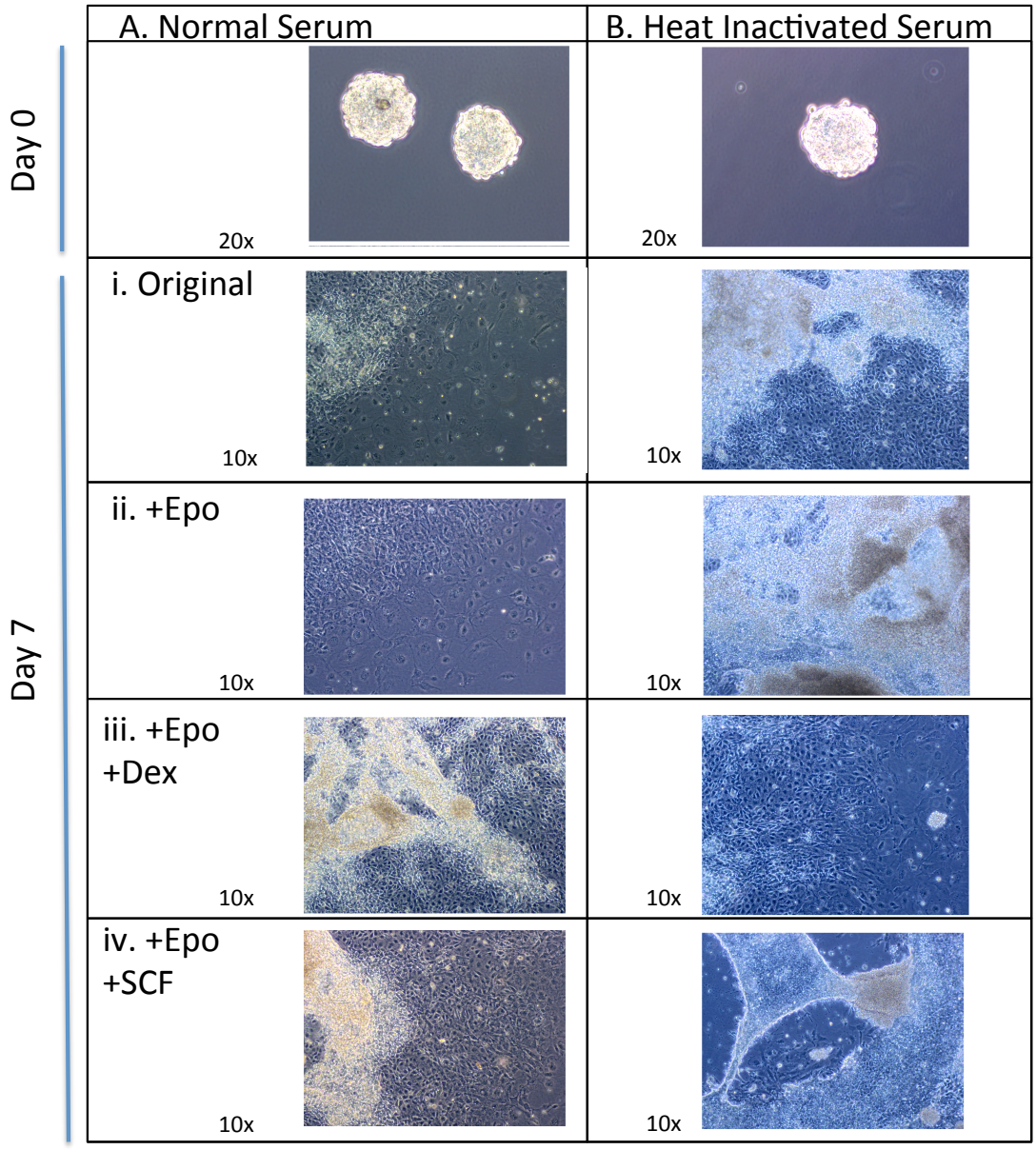


Figure 3.6 *In vitro* differentiation of WTH3 mES cells using the hanging drop method with media variants.
(A) Normal vs. (**B**) heat inactivated serum was used in combination with **(i)** original media and media with additives **(ii)** Erythropoietin, **(iii)** Erythropoietin + Dexamethasone and **(iv)** Erythropoietin + Stem Cell Factor. No red blood islands were visible in adherent cells and no red blood pellets were observed in the pelleted cell media (data not shown). Two independent biological replicates of *in vitro* differentiation using the hanging drop method with heat inactivated serum and original media were completed and gave similar results (data not shown). The images were taken with a Zeiss axiovert 25s inverted microscope at the magnification indicated in the figure.

3.2.2.3 Combined method of *in vitro* differentiation

Although the hanging drop method did not produce any visible red blood cells, the reproducibility of EB formation was very promising. This suggested an adaptation of the suspension method, which had occasionally produced red blood cells, would lend reproducibility between replicates. As such, a modified protocol, here termed the "combined method", was developed (Figure 3.7). This protocol was the same as that of the suspension method up until Day 0 when EBs were formed from a cell suspension in hanging drops of EB-Differentiation media. These EBs were harvested at Day 2 and plated in suspension in fresh EB-Differentiation media. The EBs were harvested at approximately Day 7 for downstream analysis. The combined method generated EBs of a reasonably consistent size and shape. Red blood islands were clearly visible in EBs produced using the combined method (Figure 3.8Aii).

Importantly, the combined method produced greater numbers of red blood cells compared with the suspension method, as determined by flow cytometry after staining cells for Ter-119 (Figure 3.8B), an erythroid cell surface marker (Kina et al., 2000). With the suspension method, a single $10cm^3$ dish of differentiated WTH3 cells produced only 0.07×10^5 Ter119 positive cells (from a starting number of 40,000 cells at Day 0), whereas 1.9×10^5 Ter119 positive cells were produced using the combined method (plating 200 EBs with a starting number of 60,000 cells at Day 0).

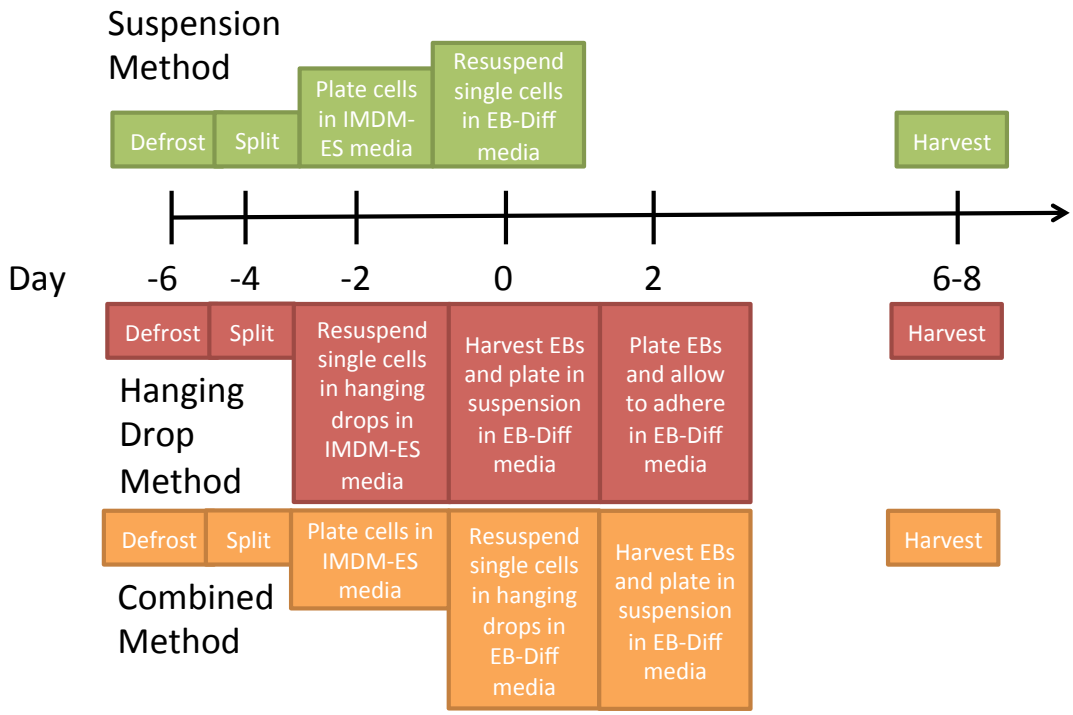
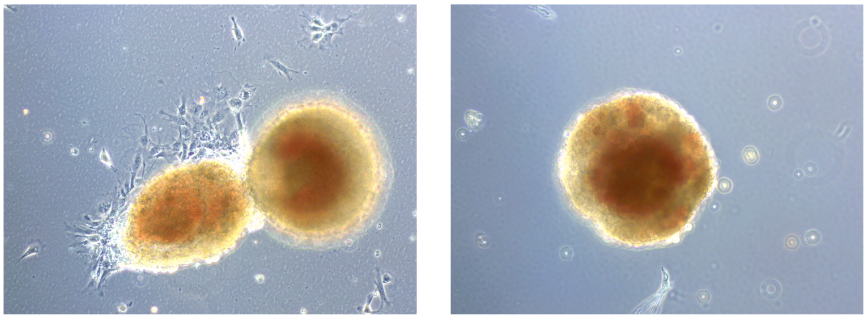


Figure 3.7 Schematic showing *in vitro* differentiation using the suspension vs. the hanging drop method. The combined method (orange) differs from the suspension method (green) and the hanging drop method (red) in the way EBs are formed (day 0 in suspension method and combined methods vs. day -2 in hanging drop method) and allowed to grow in suspension vs. adherence.

A



(i) Day 7 Suspension Method EB 10x

(ii) Day 7 Combined Method EB 10x

B

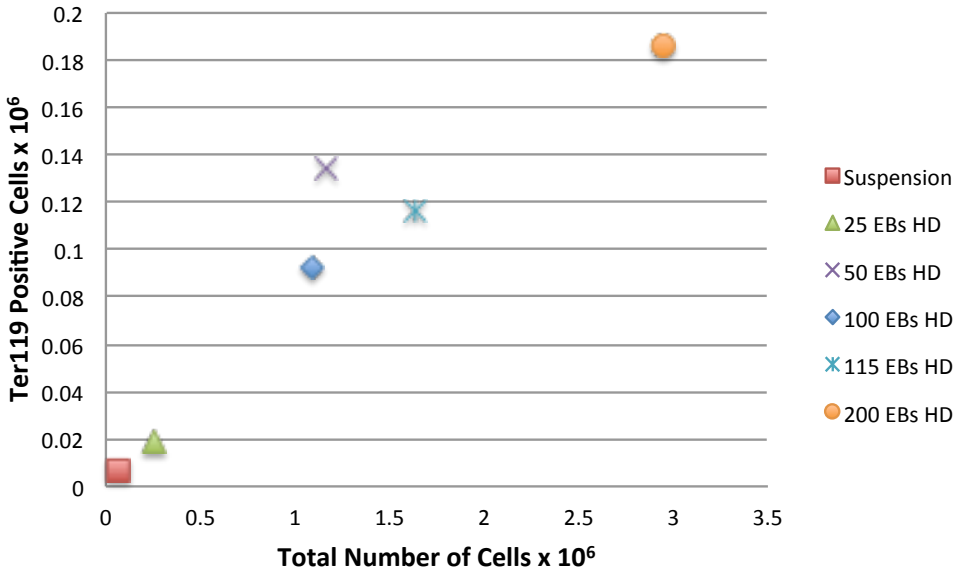


Figure 3.8 Comparison of *in vitro* differentiation using the suspension method and combined method.

A. Microscope images of EBs containing blood islands at Day 7 *in vitro* differentiation using the **(i)** suspension and the **(ii)** combined method. These images were taken with a Zeiss axiovert 25s inverted microscope at the magnification indicated in the figure.

B. Graph showing the number of Ter119 positive cells estimated by flow cytometry analysis at day 7 in vitro differentiation using the suspension method and varying numbers of hanging drop EBs (25-200 EBs HD per plate) used in the combined method.

Two independent biological replicates of *in vitro* differentiation using the combined method on WTH3 and other cell clones were performed, both with similarly high numbers of EBs containing red blood islands.

3.2.3 Molecular analysis of WTH3 mES Cells and *in vitro* differentiated cells

RNA was extracted from mouse E14 ES cells, human H9 ES cells (received as a gift from D. Hay's lab, The University of Edinburgh), humanised mouse WTH3 ES cells and humanised mouse WTH3 ES cells which had been differentiated *in vitro* towards the erythroid lineage. These RNA samples were reverse transcribed and cDNA was generated and analysed by qPCR (SYBR Green). In agreement with previously published data (De Gobbi et al., 2011), no α- or β-globin transcripts were detected in E14 mES cells, however, human α-globin was detected in human ES cells and WTH3 humanised mES cells (Figure 3.9A). No signal from mouse α- or β-globin was observed in human, humanised or mES cells (Figure 3.9B). As expected, human α-, mouse α- and mouse β- globin transcripts were all detected in WTH3 humanised mouse cells differentiated *in vitro* towards the erythroid lineage (Figure 3.9A and B). The observed expression data from these cell lines is summarised in Table 3.3.

Table 3.3 Summary of expression data from cells from ES cells and *in vitro* differentiated WTH3 cells. The expression of human α-globin is detected in humanised mouse (WTH3) and human (H9) ES cells whereas mouse α- or β-Globin is not.

	Human α	Mouse α	Mouse β
Mouse ES cell (E14)	No	No	No
Human ES cell (H9)	Yes	No	No
Humanised mouse ES cell (WTH3)	Yes	No	No
***In vitro* differentiated WTH3**	Yes	Yes	Yes

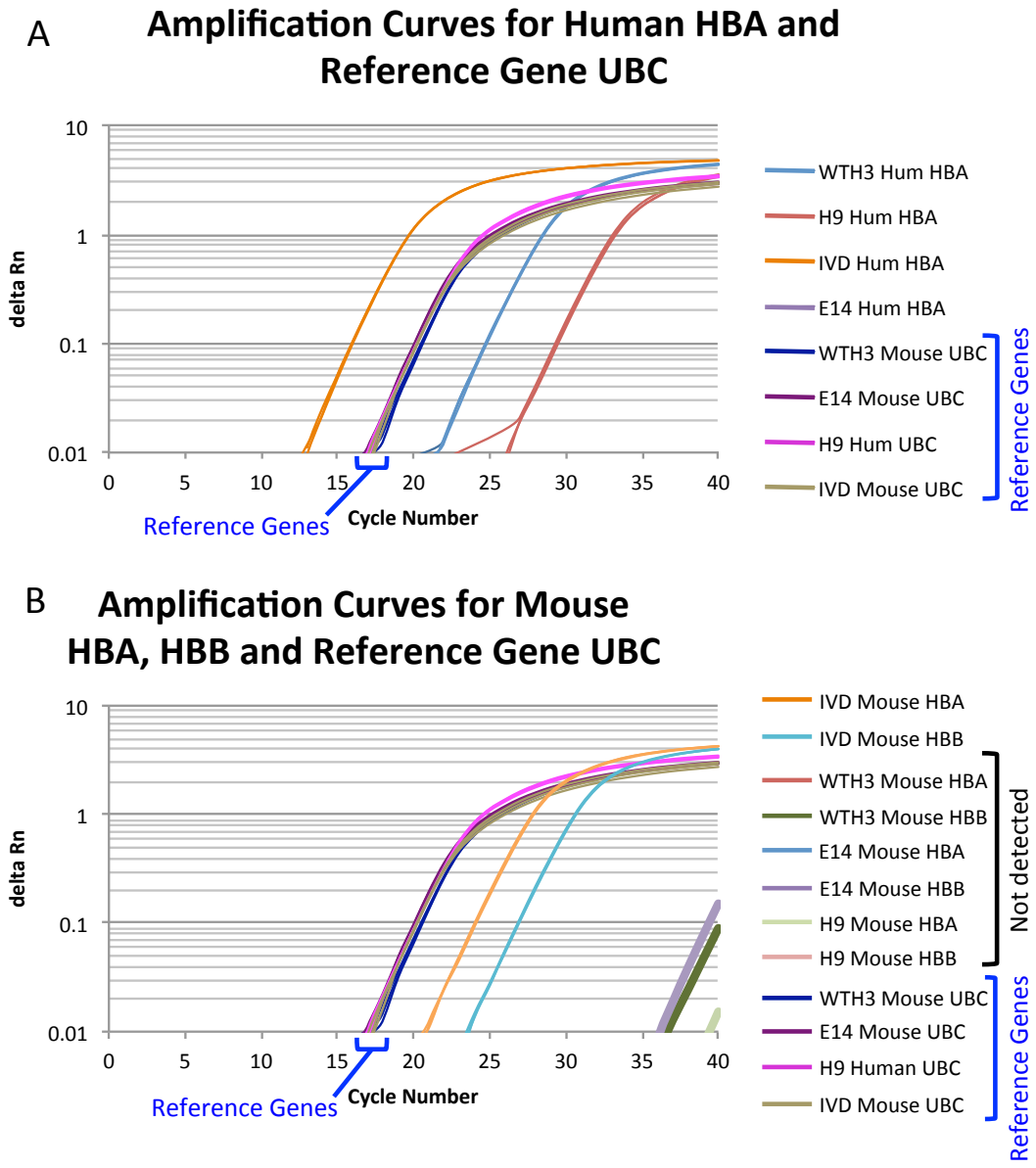


Figure 3.9 RT-qPCR amplification curves showing expression of human and mouse α-globin and mouse β-globin in WTH3 mES cells, H9 hES cells and *in vitro* differentiated WTH3 cells.
(A) Human α-globin (Hum HBA) is expressed in WTH3 humanised ES cells and H9 human ES cells. Expression of Hum HBA is detected at higher levels in WTH3 *in vitro* differentiated cells. Hum HBA expression is not detected in E14 mES cells. The reference house-keeping gene Ubiquitin C (UBC) is expressed in all samples.
(B) Mouse α-globin (Mouse HBA) and mouse β-globin (Mouse HBB) is detected in WTH3 *in vitro* differentiated cells but not in WTH3 mES cells, H9 hES cells or E14 mES cells. The reference house-keeping gene UBC is expressed in all samples.

3.3 Discussion

In this chapter, the WTH and ΔHS-40 mES cell lines were subcloned and two suitable subclones, WTH3 and ΔHS-40_20, were isolated. These mES cell subclones contain a high percentage of euploid cells (>80% for WTH3 and >50% for ΔHS-40_20) and so should be appropriate for gene targeting experiments. The ability of these cells to differentiate *in vitro* was also confirmed which suggests they are suitable lines for generating mouse chimeras and performing *in vitro* differentiation assays.

A robust *in vitro* differentiation protocol (combined method) for generating erythroid cells was optimised for use with the WTH3 mES cell line. Reproducibility issues observed here using the suspension method for generating EBs have been previously documented (Kurosawa, 2007). The use of heat inactivated serum in cell culture protocols has been traditionally used as a means of inactivating heat labile complement proteins which can negatively affect cell growth (Soltis et al., 1979). However, more recent comparisons have shown that heat inactivation often has no or a negative effect on the growth of different cell lines (Hauptman 2009). The use of heat-inactivated serum during *in vitro* differentiation had a positive effect on cell recovery at the point of harvesting and so was used in further optimisation of these protocols. The hanging drop method was able to produce reproducible EBs but did not generate any erythroid cells with visible haemoglobin expression.

The media supplements tested with the suspension method and hanging drop method (Epo, Dex, SCF) are often used, in combination with others, in re-plating two-step approaches for generating erythroid cells. The supplements did not appear to have any effect on the qualitative appearance of red blood cells in the methods tested here, although this was not quantitatively assessed and so a more subtle effect may have been achieved from the addition of these supplements. As these supplements are quite expensive and did not qualitatively appear to have an effect on the appearance of red blood cells, I decided to instead test a combined method of *in vitro* differentiation, integrating steps from both the suspension and the hanging drop methods. This combined method was capable of consistently producing higher numbers of erythroid cells than previously tested methods.

Finally, the expression of human and mouse α-globin and mouse β-globin was tested by RT-qPCR in WTH3 humanised mES cells, H9 hES cells, E14 mES cells and WTH3 cells *in vitro* differentiated to the erythroid lineage. Human α-globin was detected in both WTH3 humanised mES cells and H9 hES cells whereas mouse α- and β-globin transcripts are not detected in these cell types, nor in E14 mES cells (Figure 3.9). The humanised mES cell line is capable of reproducing expression patterns observed in the H9 hES cells and is thus a suitable model for studying human α-globin transcriptional regulation. The detection of human and mouse α-globin and mouse β-globin in the *in vitro* differentiated humanised WTH3 cells suggested that definitive erythroid cells were capable of being produced by the *in vitro* differentiation protocols used in this study.

The isolation and validation of the WTH3 and ΔHS-40_20 mES cell lines provide essential tools that the experiments in the following chapters rely on. An important aim of this thesis is the study of epigenetic regulation at the α-globin locus during development. The successful optimisation of an *in vitro* differentiation protocol to produce definitive erythroid cells was an essential precursor to future analysis of erythropoiesis.

Chapter 4 Construct Preparation for Gene Targeting

4.1 Introduction

In this chapter, I aimed to prepare a fluorescent reporter that would mark cells associated with α-globin gene activity in wild type and ΔHS-40 enhancer humanised mES cell backgrounds. Either the *HBA1* or *HBA2* genes could be fused with the coding region of a fluorescent reporter. However, as the humanised mice with a HS-40 deletion express <2% of normal α-globin mRNA levels in erythroid cells (Vernimmen et al., 2009), it was necessary to consider strategies to maximise the detection of a tagged α-globin gene. Given that the *HBA2* gene expression is 2-3 times higher than *HBA1* from approximately 8 weeks *in utero* (Albitar et al., 1992; Liebhaber et al., 1986), I focused on introducing a *2A-mCherry* knock-in at the 3` end of *HBA2* in humanised mES cells. Here, I describe the generation of a number of gene-editing tools that were used to generate knock-in mES cell lines.

Fusing genes with coding sequences that produce fluorescent proteins facilitates the visualisation of transcription in individual cells. Fluorescent reporters have been developed that have a range of excitation and emission profiles as well as varying levels of photostability. Monomeric Cherry (mCherry) was developed as a variant of the first red fluorescent protein isolated from *Discosoma sp.*, DsRed (Shaner et al., 2004). The mCherry reporter has an excitation maximum of 587 nm, is highly photostable, and is more tolerant of 5` protein fusions than its predecessor (Shaner et al., 2004). This reporter has been successfully used for a range of different applications in many cell types, including erythroid cells (van Dijk et al., 2010; Ghamari et al., 2013). For example, a mCherry knock-in mouse was used to visualise the expression of the CDK9 transcription factor by live cell imaging of a number of different tissues, including foetal liver erythroid cells (Ghamari et al., 2013). Furthermore, live imaging of β-globin transcription has been achieved using mCherry (Martins et al., 2011). The knock-in gene I aim to generate includes a 2A peptide sequence between the *HBA2* and *mCherry* genes. The 2A peptide sequence causes interrupted translation through a ribosomal skip mechanism, without affecting the rate of translation (Szymczak et al., 2004). This allows for stoichiometric expression from

2A-linked genes and will facilitate the quantification of α-globin expression in individual cells *via* mCherry detection.

Gene targeting via *recombineering technology*

Gene targeting techniques allow site-specific manipulation of a target locus to generate a deletion, mutation or insertion of a defined sequence. Two of these techniques include recombineering and CRISPR/Cas9, as previously described in detail (Chapter 1, pp. 32 and pp. 38, respectively). Recombineering techniques involve the use of bacteriophage recombineering proteins which are expressed in *E. Coli* cells and homology directed repair (HDR) mechanisms in mammalian cells. Briefly, a gene is first manipulated *in vitro,* for example in a BAC, HDR donor vectors and gene targeting vectors are generated, and then multiple gene targeting steps in ES cells are performed and chimeras are generated. This system uses recombination proteins Redα, Redβ and Gam from the λ phage which have been inserted into the bacterial genome of *E. Coli* strains (Muyrers et al., 2000). The recombinase mediated cassette exchange (RMCE), which is conducted in ES cells, relies on the use of site-specific recombinases that mediate the directed exchange of counter selective cassettes (Liu, 2013). Fragments of >100 kb have been successfully knocked-in using recombineering techniques (Wallace et al., 2007).

Gene targeting via *CRISPR/Cas9 technology*

CRISPR/Cas9 technology is used to induce double stranded breaks at a specific target sites. In the presence of an appropriately designed HDR donor vector, an insert can be introduced into a specific locus (Ran et al., 2013a). In order to prevent the digestion of the HDR donor vector, or the inserted DNA fragment after HDR, by the Cas9 nuclease, it is necessary to modify the sequence that the guide RNA targets. In many cases the modification of the donor DNA is trivial, as the guide RNA target sequences are not important for gene function or silent mutations can simply be introduced. Notably, the 3`UTR of the α-globin genes are required for their

stability (Weiss and Liebhaber, 1995). It is therefore undesirable for a random indel to be introduced after the *2A-mCherry* insertion, as this is likely to reduce the expression of the fusion gene. For this reason, I designed a strategy to introduce a previously tested 4 bp mutation into the 3`UTR of the donor plasmid at the target site of the designed guide RNA. This 3`UTR mutation does not affect the stability of *HBA2* mRNA (Weiss and Liebhaber, 1995) but does result in mismatches between the target site and the seed region of the designed guide RNA. These mismatches are likely to dramatically reduce the efficiency with which Cas9-mediated cleavage will occur (Wu et al., 2014).

Vectors to facilitate CRISPR/Cas9 mediated HDR and counter-selection of gene cassettes

The HDR vectors used in the CRISPR/Cas9 gene targeting strategy described in this chapter (Figure 4.2) were generated using Gibson assembly, an efficient PCR based cloning method (Gibson et al., 2009). Briefly, a short ~20 bp overlap is introduced into two DNA fragments by high fidelity PCR. T5 exonuclease is then used to digest the 5` end of each strand and this is counteracted by Q5 High Fidelity polymerase activity, which simultaneously replaces digested DNA. T4 DNA ligase is then used to ligate the two fragments of DNA.

CRISPR/Cas9 gene editing efficiency is extremely dependent on the genomic locus and guide RNA sequence, and also on the cell type and stage used. Although NHEJ efficiency can range from 1-90% (Bassett et al., 2013; Mandal et al., 2014; Ran et al., 2013a; Wang et al., 2013), the range of HDR efficiency is generally much lower at 0.5-30% (Ran et al., 2013a; Yang et al., 2013). The integration of a selection cassette into a HDR donor vector increases mES cell targeting screening efficiency (reviewed in Gonzalez 2016). Antibiotic selection cassettes were traditionally integrated and excised using Cre/*loxP* or FLP/*FRT* recombination technologies. A single *loxP* or *FRT* site remains after excision, which has the potential to recombine with other site-specific recombination target sites in the genome. For example, the WTH3 mES cell line used in this study contains an *FRT* and a *lox511* site within the

human α-globin locus which remain from previous targeting events (Wallace et al., 2007). The use of *piggyBac* transposase technology overcomes this problem. A counter selective cassette can be flanked by *piggyBac* inverted repeats which can be seamlessly excised in the presence of the *piggyBac* transposase without leaving a 'scar' (Yusa et al., 2011). In this chapter, I generated two HDR donor vectors, one without a selection cassette and one containing a *PGK-puromycinΔthymidine-kinase* counter-selection cassette flanked with *piggyBac* inverted repeats (*PB:PuroΔTK*). The use of the counter-selection cassette allows for the positive selection of correctly targeted mES cells based on puromycin resistance. Following confirmation of correct integration, the *piggyBac* transposase can be transiently expressed to facilitate the seamless removal of the *PB:PuroΔTK* counter-selection cassette. Clones in which the counter-selection cassette is excised can be selected by a resistance to fialuridine (FIAU) and susceptibility to puromycin.

This chapter describes the recombineering strategies designed to generate humanised mES cells with a knock-in fluorescent reporter tag, which culminated in the generation of a BAC containing a 2A-mCherry fusion. A number of CRISPR/Cas9 genome editing strategies to target this locus are also described. A guide RNA that targets the *HBA2* 3`UTR was generated and validated while HDR donor vectors (with and without selection cassettes) were produced by successive rounds of Gibson assembly. A mutation in the *HBA2* 3`UTR to disrupt the binding site of the guide RNA was included in the HDR donor vectors and the effects of this mutation on HBA2 and mCherry protein and RNA expression were tested.

4.2 Results

4.2.1 Recombineering and RMCE strategy for gene targeting at *HBA2*

In this section I describe the strategy and work done to prepare HDR donor and targeting vectors for generating a fluorescent reporter knock-in in mES cells by recombineering and RMCE. I successfully completed two rounds of recombineering to generate a BAC containing *HBA2-2A-mCherry*. However, due to the time consuming nature of recombineering, it was then decided to instead investigate CRISPR/Cas9 technology as a means of gene targeting.

The humanised mES cell line used in this project was generated using site-specific recombinase technology in combination with recombineering strategies. A BAC (BAC344L6) that was generated as part of the Human Genome Project and contained >190 kb of the human α-globin locus (Al-Hasani et al., 2004; Osoegawa et al., 2001) was used for the gene targeting strategies. As described in Chapter 1 (pp. 28) BAC344L6 was triple targeted and 117 kb of the human α-globin locus was exchanged with 85 kb of the mouse α-globin locus in mouse ES cells to generate the WTH cell line (Wallace et al., 2007). A similar gene targeting approach using a recombineering strategy was designed to knock-in a *2A-mCherry* fluorescent reporter into the subcloned WTH3 and ΔHS-40_20 mES cells at the *HBA2* gene. However, the humanised locus within these two cell lines contain *FRT*, *F3* and *lox511* sites that remain from previous gene targeting strategies (Wallace et al., 2007). The RMCE entry donor and targeting vectors, received as a gift from Dr. Andrew Smith (CRM, University of Edinburgh) also contain the site-specific recombination target sites, *loxP* and *FRT*. These sites need to be replaced with *rox* and *F5* sites, where appropriate, in order to prevent spurious recombination events when targeting the humanised mES cells (Figure 4.1).

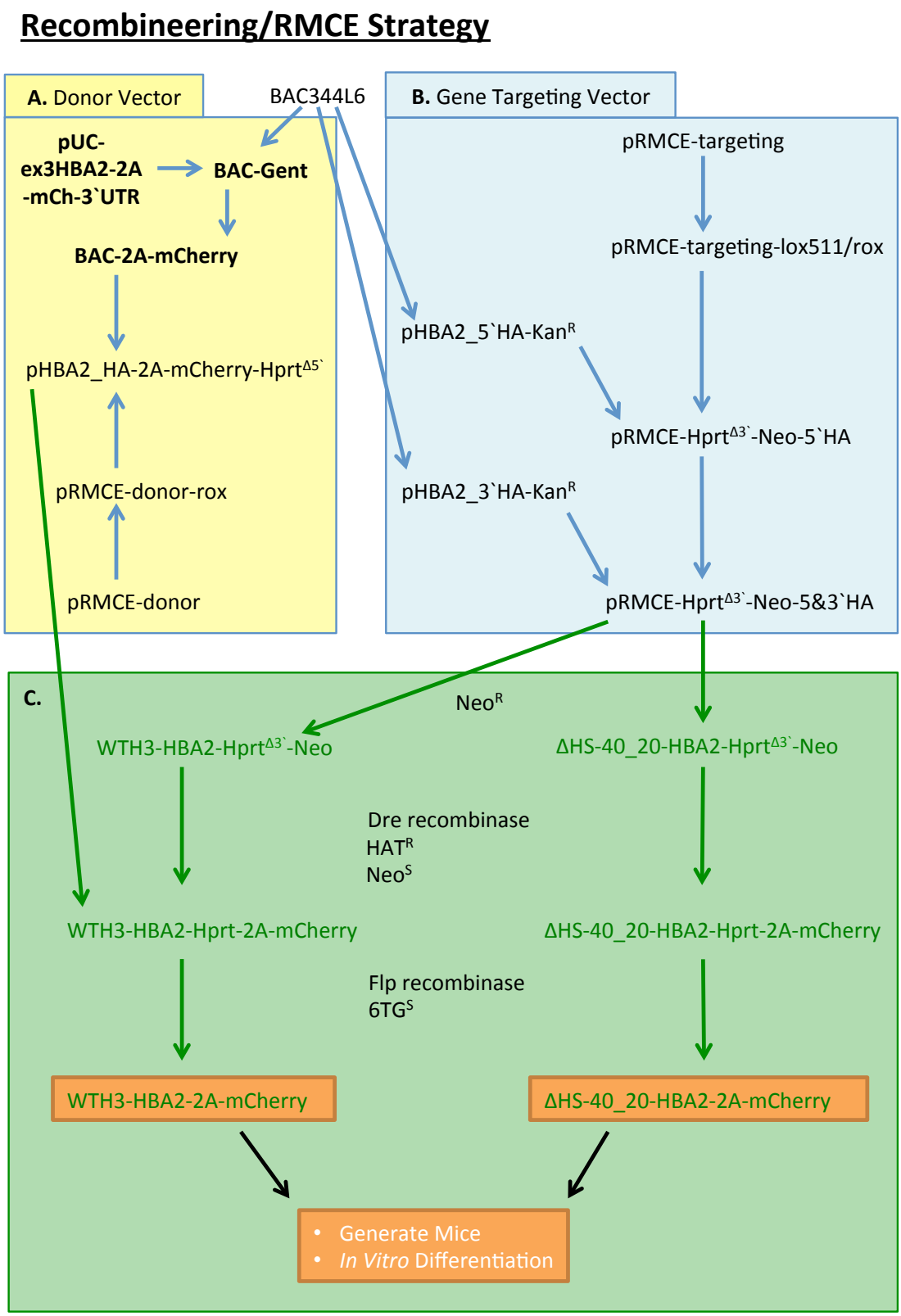


Figure 4.1 Recombineering/RMCE strategy for generating a *2A-mCherry* knock-in humanised mES cells with a wild type α-globin locus and with a deleted HS-40 enhancer. *Figure legend is continued on the next page.*

The gene recombineering/RMCE strategy outlined in Figure 4.1 involves the generation of a donor vector (Figure 4.1A). Four cloning steps in BACs are required to generate this donor vector, including three recombineering steps in BACs. In the first recombineering steps, a *rpsL-Gentamycin* (*rpsL-Gent*) cassette is to be fused to the 3` end of *HBA2* in a BAC containing the human α-globin locus, BAC344L6, to generate BAC Gent. The second recombineering step involves the exchange of the *rpsL-Gent* cassette in BAC Gent with a *2A-mCherry* fluorescent reporter cassette to generate BAC-2A-mCherry. The entry donor vector, pRMCE-donor, requires the replacement of site-specific recombination *loxP* sites with *rox* sites to generate pRMCE-donor-rox. The final donor vector (pHBA2_HA-2A-mCherry-Hprt$^{\Delta 5`}$) contains 10 kb of homology to the human *HBA2-2A-mCherry* locus and includes the 3` end of the *HPRT* minigene, upstream of the *HBA2* gene. To generate the final donor vector, the *HBA2-2A-mCherry* cassette can be retrieved from BAC-2A-mCherry into the modified entry donor vector, pRMCE-donor-rox.

A number of recombineering steps are also required in this strategy to generate the gene-targeting vector (Figure 4.1B). First, a cloning step is required to replace the site-specific recombination sites in the gene targeting entry vector, pRMCE-targeting, with *lox511* and *rox* sites to generate pRMCE-targeting-

---

Figure 4.1 *Legend continued:*
The original strategy for recombineering involves 9 cloning steps, many of which require intense screening, and 6 RMCE cell targeting steps. In this gene targeting strategy a donor vector (pHBA2_HA-2A-mCherry-HprtΔ5`) and a gene-targeting vector (pRMCE-HprtΔ3`-Neo-5&3`HA) are to be generated (**A** and **B**, respectively). RMCE must then be performed using these two vectors in the mES cell lines, WTH3 and ΔHS-40_20, to generate humanised mES cell lines with a fluorescent reporter knock-in **(C)**. The recombineering and cloning steps to be performed on BACs or plasmids are shown as blue arrows. Green arrows indicate gene-targeting steps to be performed in the mES cell lines, which are shown in green text. Steps that were completed are shown in bold text. The gene targeting strategy was designed to generate two knock-in mES cell lines (orange boxes with green text), which would then be used to generate mouse lines by blastocyst injection. Primary erythroblasts generated from mouse lines and erythroblasts generated by *in vitro* differentiation from mES cells would then be analysed.

lox511/rox. A 5` and a 3` homology arm targeting the *HBA2* locus are retrieved from BAC344L6 by recombineering and two subcloning vectors are thus generated (pHBA2_5` and 3`HA-Kan^{R}). These two homology arms are sequentially subcloned into the modified gene targeting entry vector to generate the final gene targeting vector, pRMCE-Hprt$^{\Delta 3`}$-Neo-5&3`HA. This final gene-targeting vector includes the 5` end of the *HPRT* minigene upstream of the *HBA2* gene.

Once the gene targeting and donor vectors are completed, three RMCE steps are undertaken in the WTH3 and ΔHS-40_20 mES cell lines (Figure 4.1C). Transfection of humanised mES cells with (1) the final gene-targeting vector confers neomycin resistance when integrated into the mES cell genome. Co-transfection of (2) the final donor vector with a plasmid expressing Dre recombinase into correctly targeted mES cells generates correctly targeted mES cells that are both HAT resistant and neomycin susceptible and contain a *2A-mCherry* reporter cassette at *HBA2*. A final transfection with (3) a plasmid expressing the Flp recombinase gene mediates the removal of the *Hprt* minigene and yields the final two humanised cell lines with a fluorescent reporter knock-in. These mES cell lines can then be used to produce erythroid cells *via* the generation of mice and the use of *in vitro* differentiation techniques.

4.2.1.1 Recombineering strategy to generate a BAC with a *HBA2-2A-mCherry* fusion.

As briefly described above (section 4.2.1), an *in vitro* recombineering strategy was designed to fuse *2A-mCherry* to the 3`UTR of *HBA2* in BAC344L6 (Figure 4.2). This BAC344L6-2A-mCherry (BACmCh) could then be used to generate a donor vector for RMCE in mES cells (Figure 4.1C) or potentially be used as a tool to optimize screening PCRs in an alternative gene-targeting strategy.

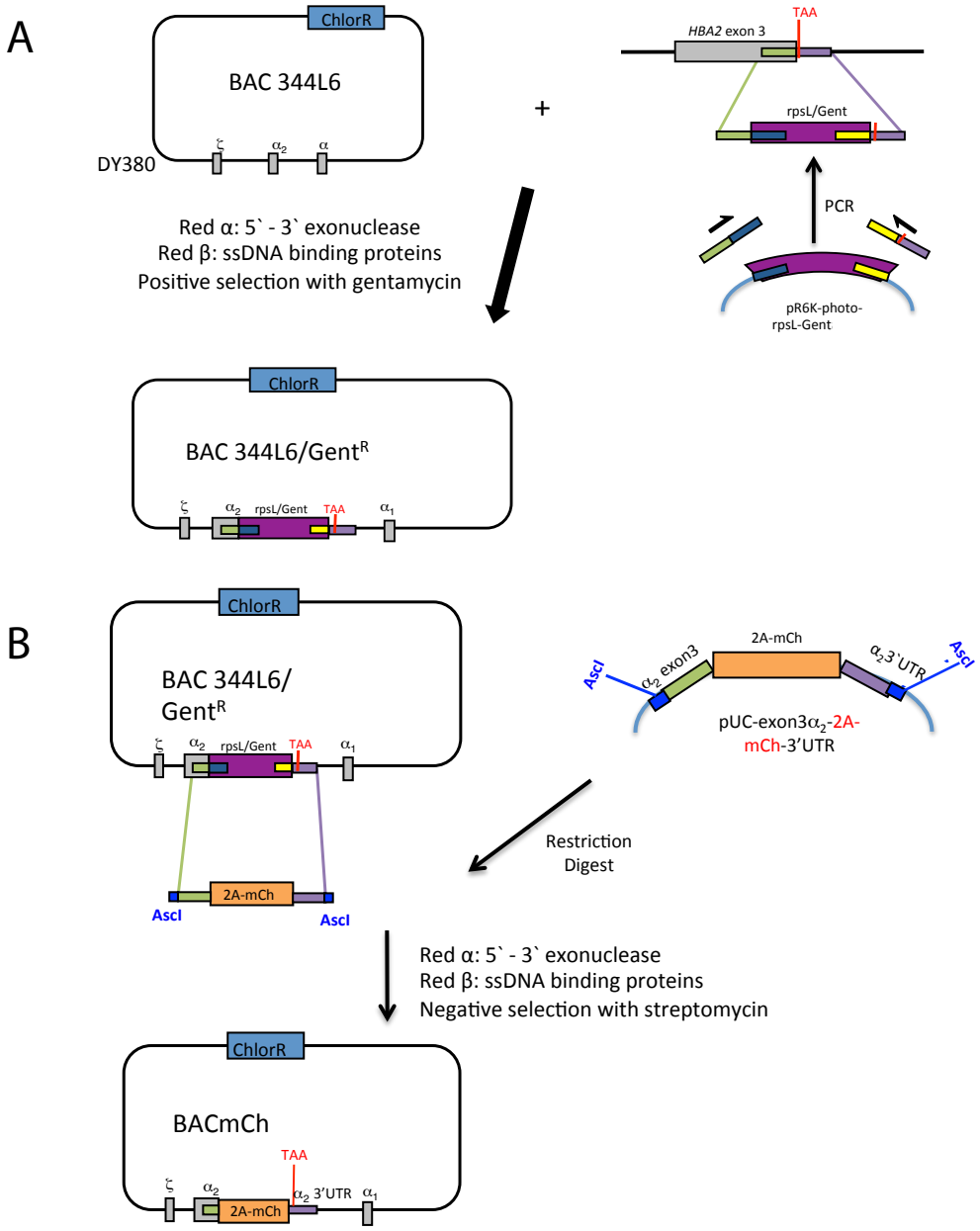


Figure 4.2 Recombineering strategy to generate a BAC with a *HBA2-2A-mCherry* fusion.

A. Strategy for the site-specific insertion of an *rpsL/Gent* selection cassette at the 3`end of *HBA2* in BAC344L6. Briefly, the *rpsL/Gent* DNA fragment was generated by PCR from a plasmid and introduced into *E. coli* containing BAC344L6 by electroporation. *E. coli* cells were induced to express recombination proteins and these cells were selected by gentamycin resistance and screened by PCR.

B. Strategy for exchange of an *rpsL/Gent* selection cassette with a *2A-mCherry* cassette in BAC344L6/Gent^R. The *2A-mCherry* (*2A-mCh*) DNA fragment was generated by restriction digestion from a plasmid and introduced by electroporation into *E. coli* containing BAC344L6/Gent^R by electroporation. *E. Coli* cells were induced to express recombination proteins and these cells were selected by streptomycin resistance and screened by PCR.

In vitro recombineering strategies are based on differential antibiotic selection. Here, an *rpsL/Gent* counter-selectable cassette was inserted at the 3`UTR of *HBA2* in BAC344L6, which was exchanged with the *2A-mCherry* cassette (Figure 4.2). Initial selection of BAC clones containing the *rpsL/Gent* cassette (BAC Gent clones) was achieved by screening for clones resistant to gentamycin and sensitive to streptomycin (Figure 4.2A). Subsequently, clones containing the *2A-mCherry* cassette were identified based on their resistance to streptomycin and sensitivity to gentamycin (Figure 4.2B).

4.2.1.2 Generation of BAC Gent

The *rpsL/Gent* fragment was generated by PCR from a plasmid pR6K-photo-rpsL-Gent. The PCR product included a 5` 50 bp homology arm to the last 50 bp of *HBA2* exon 3, and a 3` 50 bp homology arm to the first 50 bp of the *HBA2* 3`UTR (excluding the stop codon). The PCR product was introduced by electroporation to DY380 cells containing BAC344L6 that were induced to express recombination proteins. Electroporated cells were expanded on LB/Agar with 3-4 μg/ml gentamycin selection. Approximately 270 surviving colonies were re-plated on fresh LB/Agar plates with 3 μg/ml gentamycin selection. Of these, 24 clones were randomly selected and expanded in liquid LB with 3 μg/ml gentamycin selection and BAC mini preps were performed on these cultures.

Screening of selected clones for correct integration of the *rpsL/Gent* cassette was performed by PCR with the orange primers depicted in Figure 4.3A. Five PCR positive clones (BAC Gent A-E) and a PCR negative clone (BAC Gent F) were identified (Figure 4.3B). These clone were then screened by Southern blot for correct integration of *rpsL/Gent* using a *Gent* probe (Figure 4.3C). Correct insertion of *rpsL/Gent* was confirmed in BAC Gent A-D by Southern blot analysis, however BAC Gent E and F appeared to have doublets suggesting either incomplete restriction enzyme digestion (this would give bands of 4.4 kb for *Hind*III and 1.5 kb for *Hpa*I) or additional random integration events. DY380 cells containing BAC Gent A-F were also plated on LB/Agar containing 75 μg/ml streptomycin to screen for susceptibility

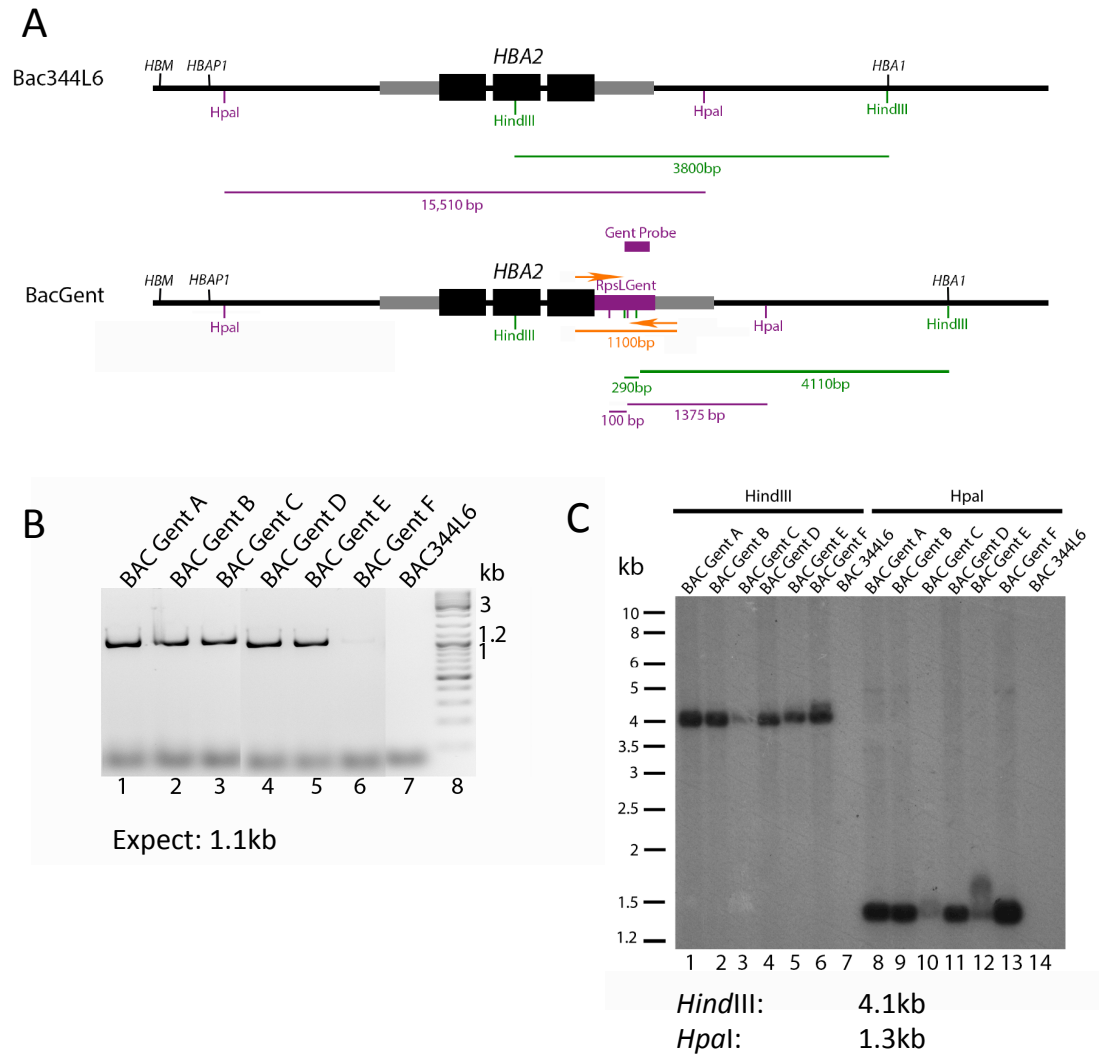


Figure 4.3 Correct fusion of *rpsL/Gent* to the last codon of *HBA2* in BAC344L6.
A. Schematic of primers, probes and restriction sites used in screening correct integration of *rpsL/Gent*. *Hpa*I and *Hind*III restriction sites using in Southern blot screening are indicated in magenta and green, respectively.
B. Composite gel image showing PCR screening of BAC Gent clones A-F. Orange arrows in **(A)** indicate primer pair used, which produce a product of 1.1 kb following correct insertion of the *rpsL/Gent* cassette and no product without insertion. Clones BAC Gent A–E (lanes 1-5) all contained the correct 1.1 kb product. BAC344L6 (lane 7) gave no product, as expected.
C. Southern blot screening of BAC Gent clones A-F using a probe for gentamycin (magenta box in A). Digestion with *Hind*III gave a single expected band size of 4.1 kb (observed in lanes 1-5) and digestion with *Hpa*I gave an expected band size of 1.3 kb (observed in lanes 8-11 and lane 13). Apparent doublets are expected to be a result of incomplete digestion of the BAC DNA (lanes 6 and 12). BAC344L6 gave no apparent band (lane 7 and 14), as expected.

to this antibiotic conferred by a functional *rpsL* gene. BAC Gent C and E were the only two clones that showed no growth on LB/Agar plates or LB liquid cultures supplemented with 75 μg/ml streptomycin. As a result of this streptomycin screen it was decided to pursue further cloning steps with BAC Gent C and E.

4.2.1.3 Generation of BAC-2A-mCherry

The donor plasmid used to insert the *2A-mCherry* cassette directly in between the final codon and the stop codon of *HBA2* was designed to contain 50 bp homology arms to *HBA2*, the 'self-cleaving' 2A peptide and the *mCherry* fluorescent reporter gene (5`-exon 3 *HBA2* - *2A-mCherry* - 3`UTR of *HBA2* -3`). This cassette was synthesized and cloned into a pUC19 vector by GenScript (pUC_ex3HBA2_2A-mCherry-3`UTR). The *2A-mCherry* fragment was generated by restriction digestion with *Asc*I. This fragment was introduced by electroporation to DY380 cells containing BAC Gent C and E that were induced to express recombination proteins. Electroporated cells were expanded on LB/Agar with 100 μg/ml streptomycin selection.

Approximately 400 clones survived from the BAC Gent C electroporation and 130 clones survived from the BAC Gent E electroporation. A random selection of 50 individual clones from each of these plates were duplicate plated on LB/Agar containing either 100 μg/ml streptomycin or 3 μg/ml gentamycin selection. BACs were isolated from 24 clones that grew on streptomycin plates but not on gentamycin plates. These clones were screened by PCR for correct integration of the *2A-mCherry* cassette using the orange primers in Figure 4.4A. BACmCh5 (a derivative of BAC Gent C) was the only positive clone identified by PCR screen (Figure 4.4B). The correct replacement of the *rpsL/Gent* cassette with the *2A-mCherry* cassette was confirmed in BACmCh5 by Southern blot using a mCherry probe (Figure 4.4C).

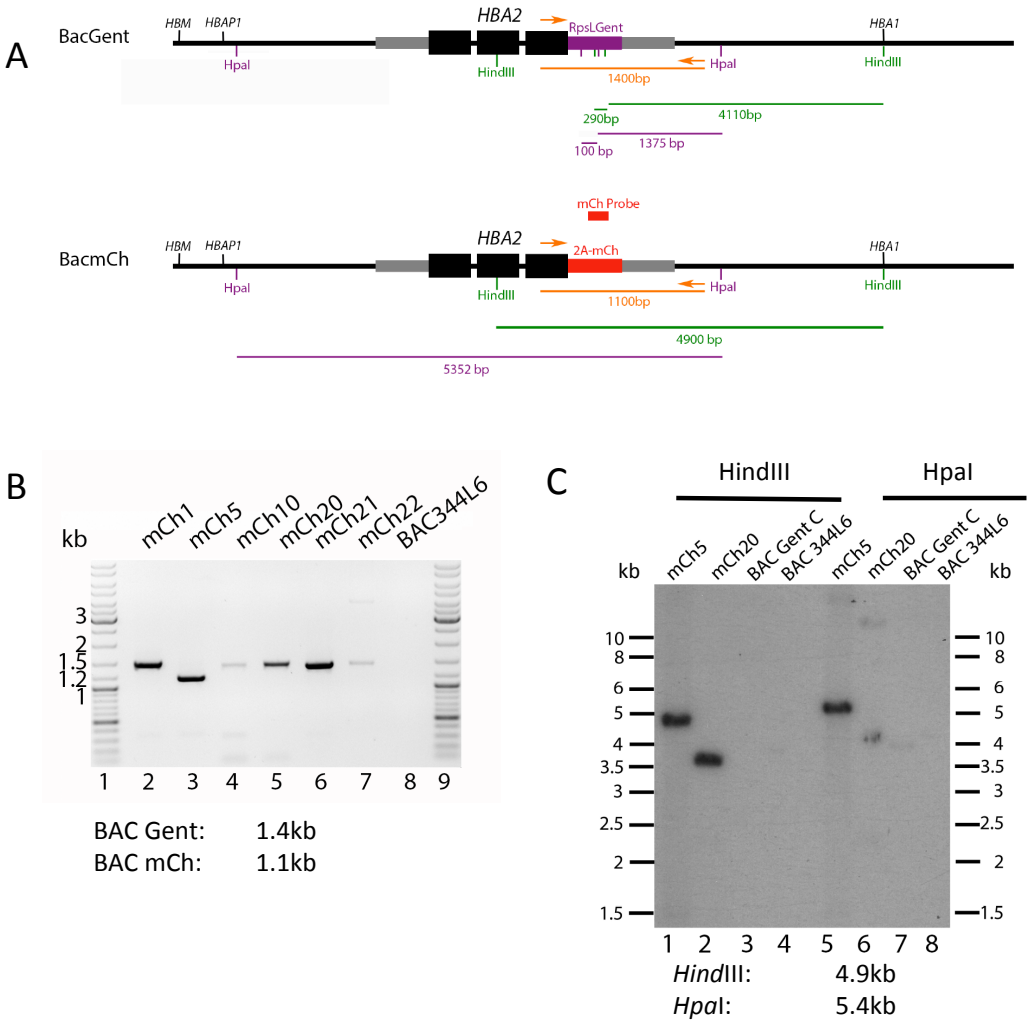


Figure 4.4 Exchange of *2A-mCherry* with *rpsL/Gent* in BAC Gent-C.
A. Schematic of primers, probes and restriction sites used in screening for correct integration of the *2A-mCherry* (2A-mCh) cassette. The *Hpa*I and *Hind*III restriction sites used in Southern blot screening are indicated in magenta and green, respectively.
B. Representative gel image showing PCR screening of BAC-mCh clones. The orange arrows in A indicate the primers used in screening that were expected to give a PCR product of 1.1 kb following exchange of *rpsL/Gent* with *2A-mCherry*. Only the mCh5 clone appeared to be the correct size (lane 3). Bands of 1.4 kb (observed in lanes 2, 4-7) suggested the *rpsL/Gent* cassette was still present at *HBA2*. The negative control BAC344L6 (lane 8) showed no amplification.
C. Southern blot screening of BAC-mCh clones using a probe for *2A-mCherry* (red box labeled 'mCh' in A). Digestion with *Hind*III gave an expected band size of 4.9 kb (observed for clone mCh5, lane 1) and digestion with *Hpa*I gave an expected band size of 5.4 kb (observed for mCh5, lane 5). The negative controls BAC Gent C and BAC344L6 lanes showed no apparent bands (lanes 3, 4, 7 and 8), as expected. Bands of lower sizes for clone mCh20 (lanes 2 and 6) indicated 2A-mCherry may have been inserted at an unknown site other than *HBA2*.

4.2.2 CRISPR/Cas9 strategy for gene targeting at *HBA2*

A second gene targeting strategy was designed to knock-in the *2A-mCherry* reporter cassette in humanised mES cell lines at the *HBA2* gene using CRISPR/Cas9 technology. This strategy required the generation of a homology directed repair (HDR) donor vector and CRISPR/Cas9 targeting vector (Figure 4.5A and B). Donor vectors were generated with relatively shorter homology arms (0.8-1.3 kb) than those required in the previously described recombineering strategy (2-5 kb). Donor vectors used in this strategy were generated by Gibson assembly. Briefly, an 800 bp 3` and a 5` homology arm to the *HBA2* locus were sequentially added from the BACmCh5 template to the donor vector pUC-ex3HBA2-2A-mCh-3`UTR that was used to generate BACmCh5. A further 500 bp was added to the 3` homology arm to generate the final HDR donor vector without a selection cassette, pAB017.

The next HDR donor vector that was generated by Gibson assembly (pAB018) contained a *PB:PuroΔTK* selection cassette. The 3` homology arm extension in pAB017 was designed such that when *PB:PuroΔTK* was inserted at an appropriate *TTAA* site downstream of the *HBA2* stop codon, an 800 bp homology arm to the *HBA2* locus would be present downstream of the counter-selection cassette. The final HDR donor vector (pAB027) was generated by modifying pAB018 to include a mutated *HBA2* 3`UTR.

To generate the targeting plasmid a single guide RNA, sgRNA-A, was designed to target the 3`UTR of *HBA2*. The sgRNA-A sequence was then cloned into a plasmid expressing Cas9 to generate the final targeting vector, pCas9-sgRNA-A. A cut at the endogenous *HBA2* locus, mediated by pCas9-sgRNA-A transfection, was validated by T7 Endonuclease I assay (Figure 4.5B). It was also important to confirm that the 3`UTR mutation included in HDR vector pAB027 did not negatively effect the expression of *HBA2-2A-mCherry in vitro*. To this end, two expression vectors were cloned which included the *HBA2-2A-mCherry* cassette and either a wild type or mutated 3`UTR. The expression of these two constructs was compared by Western blot and RT-qPCR (Figure 4.5C).

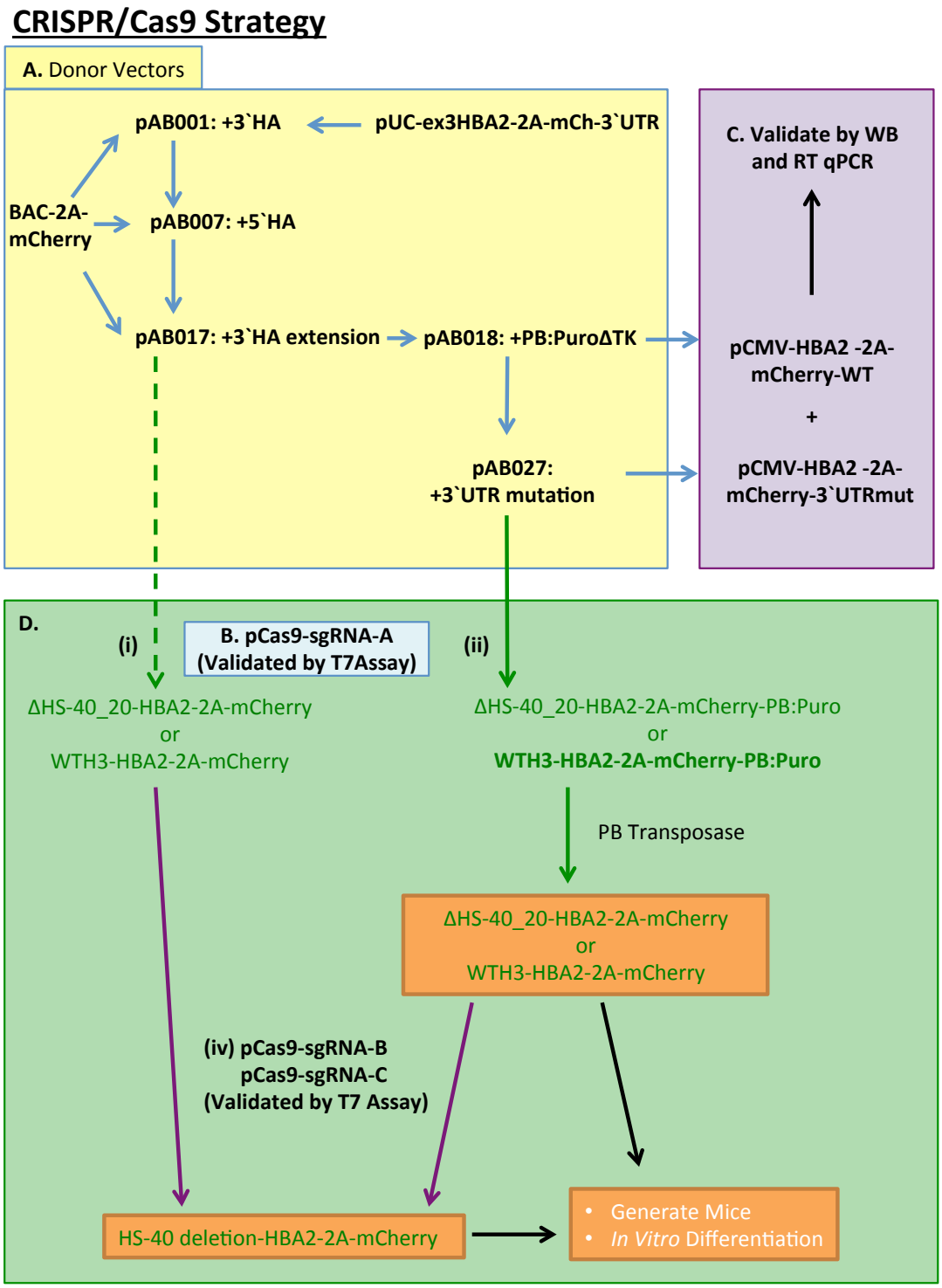


Figure 4.5 CRISPR/Cas9 strategy for generating a *2A-mCherry* knock-in humanised mES cells with a wild type α-globin locus and with a deleted HS-40 enhancer. The generation of HDR donor vectors for use in the CRISPR/Cas9 targeting strategy involved 5 cloning steps, many of which involved Gibson assembly, a highly efficient method of cloning, and 2-5 cell targeting steps, depending on which strategy was employed. *Figure legend is continued on the next page.*

Once the HDR donor and targeting vectors were completed, a number of different gene targeting strategies could be employed in the WTH3 and ΔHS-40 mES cell lines (Figure 4.5D). In a gene targeting strategy using a HDR donor vector without selection cassettes (Figure 4.5Di), WTH3 or ΔHS-40_20 mES cells are co-transfected with the targeting vector pCas9-sgRNA-A and the HDR donor vector pAB017. This would generate two humanised mES cell lines with a *HBA2-2A-mCherry* fusion that are ready for downstream analysis. Using a HDR donor vector with a selection cassette (Figure 4.5Dii) would instead yield two puromycin resistant mES cell lines, with less screening. These would need to be transfected again with a plasmid expressing a *PiggyBac* transposase to remove the selection cassette from the final FIAU resistant and puromycin sensitive mES cell lines. An alternative strategy for generating a humanised mES cell line with a deleted HS-40 enhancer was also designed (Figure 4.5Div). A pair of guide RNAs flanking the HS-40 enhancer were designed, cloned and validated by T7 Endonuclease I assay (data not shown). These could potentially be used to remove the HS-40 enhancer from the WTH3-2A-mCherry knock-in line. The final knock-in mES cell lines would then be ready for *in vitro* differentiation assays and the generation of mice.

Figure 4.5 *Legend continued:*
A number of donor vectors were cloned **(A),** as was a primary CRISPR/Cas9 targeting vector **(B)**. Two expression vectors were also cloned to determine the effect of the 3`UTR mutation on HBA2-2A-mCherry expression **(C)**. A number of different gene targeting strategies were designed **(D)**. Cloning steps performed on plasmids are shown as blue arrows. Green arrows indicate the gene-targeting steps performed in mES cell lines, which are shown in green text. The steps that were completed are shown in bold text. Dashed green arrows indicate a cell-targeting route, which was attempted but was unsuccessful in generating PCR positive clones. Purple arrows indicate an alternative route for generating the ΔHS-40-2A-mCherry knock-in cell line. The gene-targeting strategy was designed to generate two knock-in mES cell lines (orange boxes with green text) that would then be used to generate mouse lines by blastocyst injection and erythroblasts by *in vitro* differentiation from mouse ES cells.

4.2.2.1 Generation of a HDR vector without selection (pAB017)

A HDR donor vector (pAB017) was designed to knock-in a *2A-mCherry* cassette at the *HBA2* gene without the use of a selection cassette. This HDR vector was generated by sequential Gibson assembly cloning steps. Briefly, an 800 bp 5` homology arm and a 1300 bp 3` homology arm were sequentially cloned by Gibson assembly into a vector (pUC-ex3HBA2-2A-mCh-3`UTR) which contained the *2A-mCherry* reporter cassette flanked by a 50 bp 5` and 3` homology arm.

First, the 800 bp 3` homology arm was amplified by PCR from BACmCh5. Primers were designed to amplify an insert (3` arm) including the entire *HBA2* 3`UTR. The insert primers also included a ~20 bp 5` overhang homologous to the last 20 bp of *mCherry,* including the stop codon, and a ~20 bp 3` overhang homologous to the plasmid backbone in pUC-ex3HBA2-2A-mCh-3`UTR (Figure 4.6A; synthesised by GenScript) following the *HBA2* 3`UTR. A second set of primers were designed to amplify the entire vector, pUC-ex3HBA2-2A-mCh-3`UTR, excluding the *HBA2* 3`UTR segment. The vector and insert fragments were cloned into one plasmid (pAB001, Figure 4.6B) using Gibson assembly. A random selection of 10 plasmid clones were screened by restriction digestion with *Asc*I and 6 clones were determined to be have the correct restriction digestion pattern (Figure 4.6D). These clones were Sanger sequenced (data not shown) and clone pAB001.1 showed seamless integration of the 800 bp 3` homology arm.

Next, the 800 bp 5` homology arm was amplified by PCR from BACmCh5. Insert primers were designed to amplify an insert (5` arm) including the entire *HBA2* coding sequence and promoter. The insert primers also included a ~20 bp 5` overhang homologous to the plasmid backbone in pAB001 preceding the third *HBA2* exon and a 3` overhang homologous to the first 20 bp of *2A-mCherry*. The vector primers were designed to amplify the entire vector pAB001 excluding the *HBA2* exon 3 segment. The vector and insert fragments were cloned into one plasmid (pAB007, Figure 4.6C) using Gibson assembly. A selection of 5 plasmid clones were

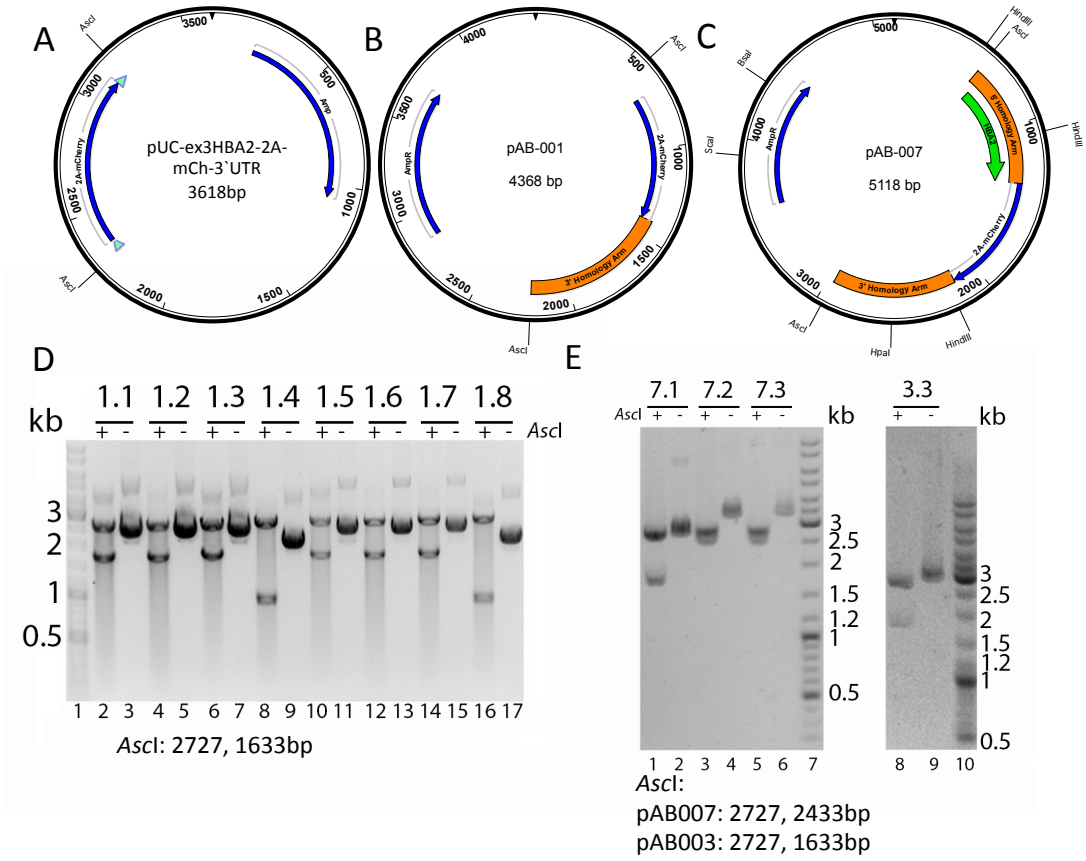


Figure 4.6 Insertion of homology arms into HDR subcloning targeting vector by Gibson assembly.

A. Plasmid map of pUC-ex3HBA2-2A-mCherry-3`UTR.

B. Plasmid map of pAB001 (pUC-ex3HBA2-2A-mCherry-3`UTR with additional 3`homology arm).

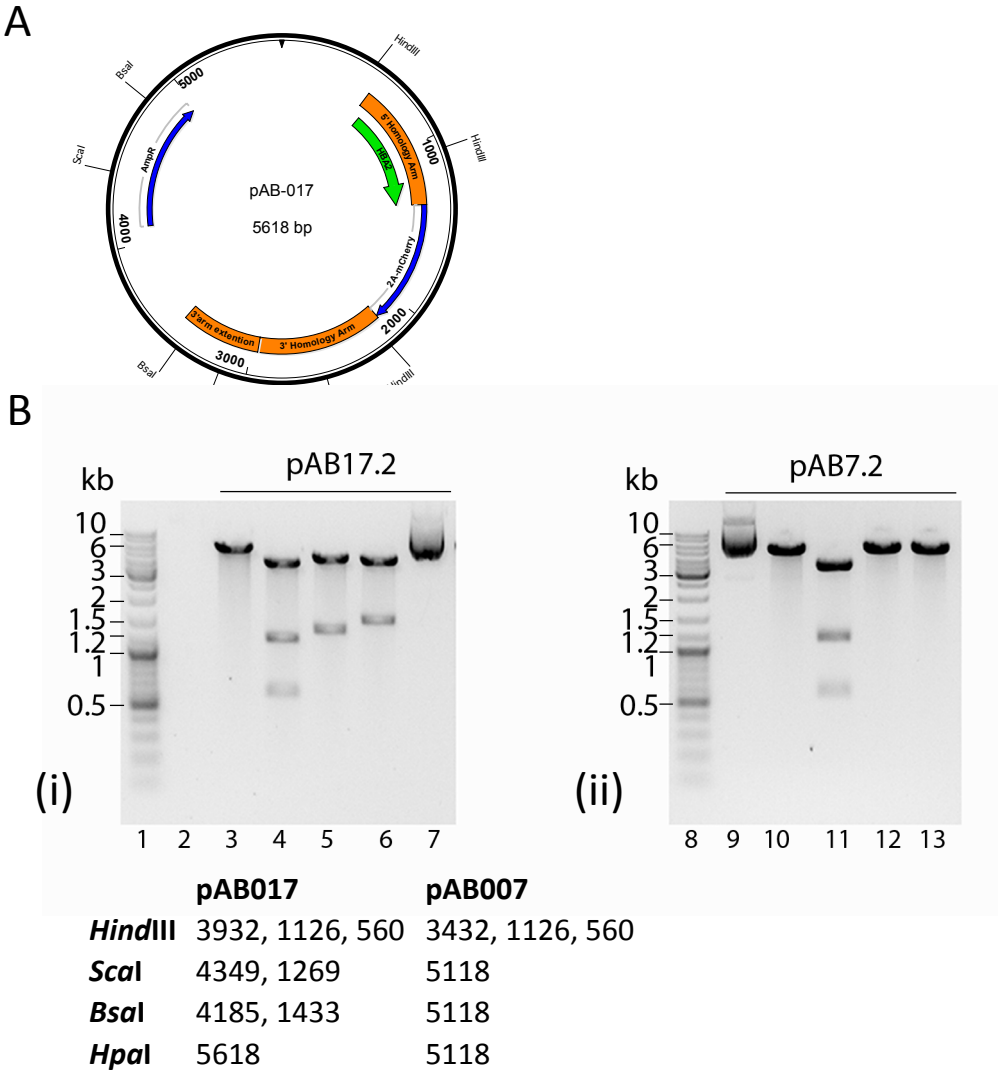
C. Plasmid map of pAB007 (pAB001 with additional 5`homology arm).

D. Representative gel image from restriction digest screen of pAB001 Gibson assembly clones. *Asc*I sites flank the entire *2A-mCherry* insert in pUC-ex3HBA2-2A-mCherry-3`UTR. Digesting with *Asc*I gives fragments of 2727 bp + 883 bp without the insert and 2727 bp +1633 bp with the addition of an 800 bp 3` homology arm. Lanes 2, 4, 6, 10, 12 and 14 contained positive clones and were sent for Sanger sequence confirmation. Clone 3.3 (not shown) was confirmed by Sanger sequencing and used in further cloning experiments.

E. Representative gel image from restriction digest screen of pAB007 Gibson assembly clones. Digestion with *Asc*I gives fragments of 2727 bp + 1633 bp without the insert (3`arm only) and 2727 bp +2433 bp with the addition of an 800 bp 5` homology arm. Clones 7.2 and 7.3 were positive and sent for sequence confirmation. The negative control, pAB-003.3 from which 5`arm clones were derived, showed the expected band sizes without insert. Clone 7.2 was used in subsequent experiments.

screened by restriction digestion with *Asc*I and 3 clones were determined to have the correct restriction digestion pattern (Figure 4.6E). These clones were analysed by Sanger sequencing (data not shown), which indicated the seamless integration of the 5` homology arm in clone pAB007.2.

Finally, a 500 bp 3` extension to the 3` homology arm was amplified by PCR from BACmCh5. The insert primers were designed to include a ~20 bp 5` overhang homologous to the last 20 bp of the 3`arm in pAB007 and a 3` overhang homologous to 20 bp of the backbone in pAB007 following the 3`arm. The vector primers were designed to amplify the entire vector pAB007. The vector and insert fragments were cloned into one plasmid (pAB017, Figure 4.8A) using Gibson assembly. A selection of 13 plasmid clones were screened by PCR and 7 clones were deemed as positive (data not shown). One of these PCR positive clones, pAB017.2 was screened by restriction digestion with multiple restriction enzymes and was determined to be correct (Figure 4.8B). This clone was analysed by Sanger sequencing (data not shown) and showed seamless integration of the 3` homology arm extension.



	pAB017	pAB007
***Hind*III**	3932, 1126, 560	3432, 1126, 560
***Sca*I**	4349, 1269	5118
***Bsa*I**	4185, 1433	5118
***Hpa*I**	5618	5118

Figure 4.7 Generation of HDR vector pAB017 by Gibson assembly.
A. Plasmid map of pAB17 (pAB007+3`arm extension)
B. The clone pAB17.2 was identified as a positive clone following Gibson assembly mediated insertion of a 3`arm extension. The plasmid identity was confirmed by restriction digestion with *Hind*III (lanes 4 and 11), *Sca*I (lanes 5 and 12), *Bsa*I (lanes 6 and 13) and *Hpa*I (lanes 7 and 10). Restriction digestion reactions of pAB17.2 (i) and the template for pAB17.2 cloning, pAB7.2 (ii), were run alongside uncut plasmid DNA (lanes 3 and 9). The expected fragment sizes are described in the figure.

4.2.2.2 Generation of a HDR vector with selection (pAB027)

To generate a HDR donor vector, which included a counter-selective cassette, the *PB:PGK:PuroΔTK* selection cassette was cloned into HDR vector pAB017 by Gibson assembly to generate pAB018 (Figure 4.8A). Firstly, the 3236 bp *PB:PGK:PuroΔTK* selection cassette was amplified by PCR from the vector pMCS-AAT-PB:PGKpuro-delta-tk, received from the Sanger Institute (Yusa et al., 2011). The insert primers were designed to amplify the selection cassette and to include a ~20 bp 5` and a 3` overhang with homology to a site within the 3` homology arm of pAB017. These primers were designed to clone the selection cassette directly downstream of a point in the 3`arm where a native *TTAA* site is present. The primers were also designed to introduce a novel *TTAA* site directly downstream of the selection cassette. Following transposase mediated removal of the selection cassette only one *TTAA* site would remain at the locus ensuring seamless excision. Vector primers were designed to amplify the entire vector pAB017. The vector and insert fragments were cloned into one plasmid (pAB018, Figure 4.8A) using Gibson Assembly. A selection of 23 plasmid clones were screened by PCR and 7 clones were determined to be correctly cloned. All of these correct plasmid clones were screened by restriction digestion with the enzyme *Hind*III and all of these clones were determined to be correct (Figure 4.8B). These clones were analysed by Sanger sequencing (data not shown) and clone pAB018.11 showed seamless integration of the *PB:PGK:PuroΔTK* selection cassette.

Finally, a ~720 bp dsDNA fragment was generated (GeneBlock, IDT) to include a 4 nt mutation in the *HBA2* 3`UTR at the sgRNA-A target site (see section 4.2.2.4, pp. 123 for details). The GeneBlock insert and vector pAB018 were digested with *Sbf*I and cloned using Gibson Assembly to generate pAB027. Correct cloning was confirmed by Sanger sequencing (data not shown). Clone pAB027.4 was used in future experiments.

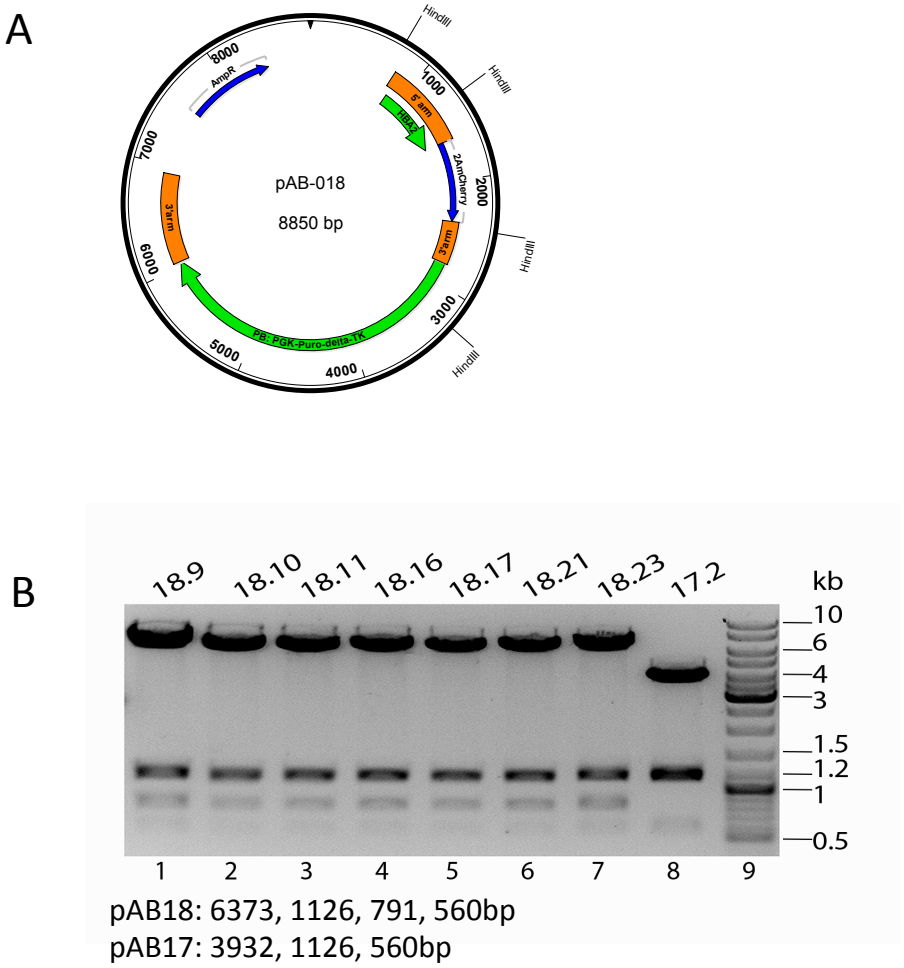


Figure 4.8 Generation of HDR vector with a selection cassette by Gibson assembly.
A. Plasmid map of pAB018 (pAB017 + PB:PGK:PuroΔTK selection cassette).
B. Screen of pAB18 cloning. Clones were screened by PCR initially and 7 plasmid clones, which were positive from PCR screens (data not shown), were digested with *Hind*III. All clones screened gave correct digestion pattern (lanes 1-7). The negative control pAB7.2 (lane 8) does not contain the selection cassette. The expected band sizes following digestion with *Hind*III are shown in the figure. Clone pAB018.11 was used in future experiments.

4.2.2.3 Generation and validation of CRISPR/Cas9 vector targeting HBA2

To fuse a *2A-mCherry* cassette at the final translated codon of *HBA2* using CRISPR/Cas9, a guide RNA was required to target the final exon or the 3`UTR of *HBA2*. In humanised mES cells the two adult α-globin genes, *HBA2* and the downstream *HBA1* are present. The coding and 3`UTR sequences of *HBA2* and *HBA1* were compared using Clustal Omega (McWilliam et al., 2013). The entire coding sequences of the *HBA2* and *HBA1* genes are identical (data not shown) however the sequences of the two genes diverge at the 3`UTR (Figure 4.9A). Using an online tool for designing guide RNAs (crispr.mit.edu) a suitable PAM and guide RNA seed sequence was identified 36 nt away from the last coding nucleotide of *HBA2*, in a region where the *HBA2* and *HBA1* sequences have diverged (shown in purple in Figure 4.9A). A single guide RNA (sgRNA-A) containing the 20 bp sequence preceding the identified PAM was cloned into an expression vector containing human codon optimised SpCas9 and a GFP fluorescent cassette (pSpCas9-2A-EGFP (PX458), (Ran et al., 2013a), received as a gift from Dr. P. Hohenstein, University of Edinburgh, Figure 4.9B). SgRNA-A targeted the *HBA2* 3`UTR specifically and contained 5 mismatches for the *HBA1* 3`UTR and 4 mismatches for the *HBA2* 3`UTR mutation included in HDR donor vector pAB027 (Figure 4.9A).

The plasmid pCas9-sgRNA-A and a vehicle control were transfected into WTH3 mES cells and positively transfected cells were isolated by FACS. GDNA was isolated from positively transfected cells and a T7 Endonuclease I assay was performed on the DNA. A band corresponding to the expected digestion pattern was observed in T7 digested cells transfected with pCas9-sgRNA-A (black arrowhead, Figure 4.9D). The indel efficiency of sgRNA-A was estimated to be 22%.

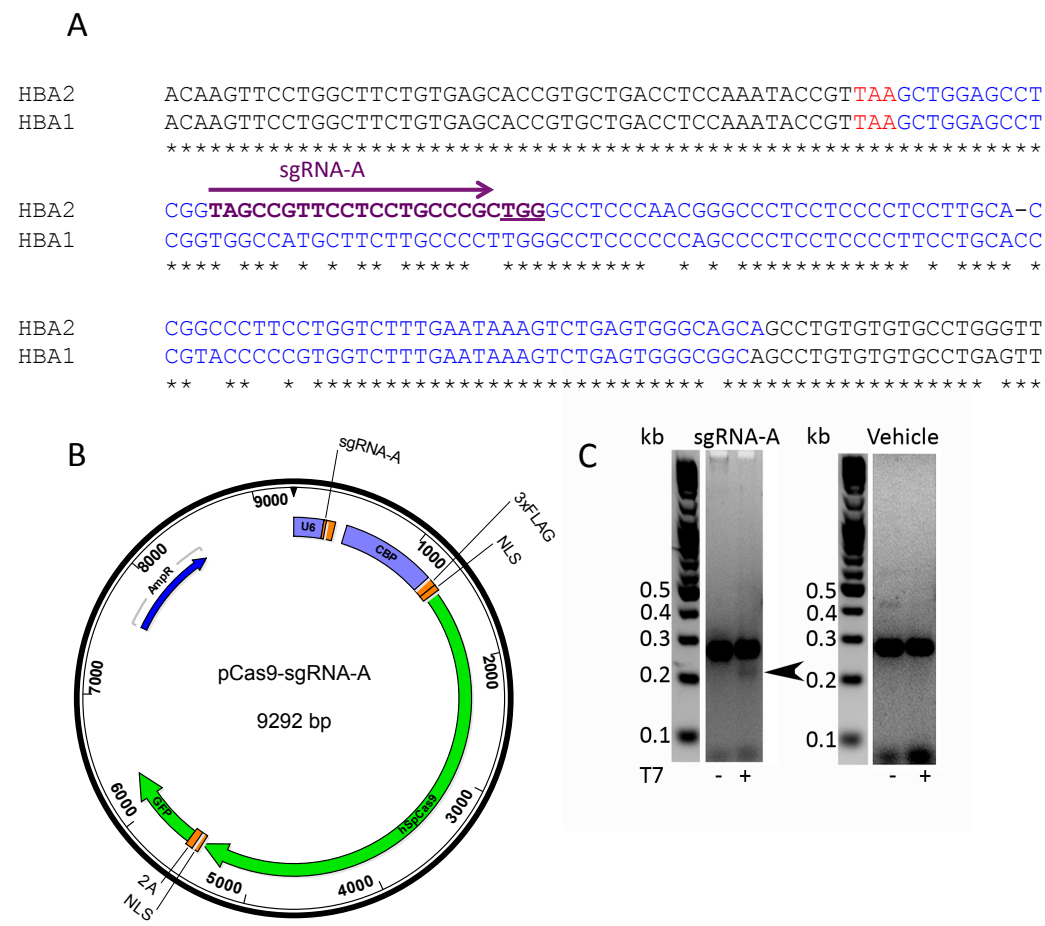


Figure 4.9 Validation of sgRNA targeting *HBA2* 3`UTR.
A. The coding and 3` UTR nucleotide sequence of human *HBA1* and *HBA2* were compared using the webtool Clustal Omega (McWilliam et al., 2013). The TAA stop codon for both genes are shown in red and the 3`UTR is shown in blue. Identical nucleotides are indicated "*" underneath the nucleotide sequence. A guide RNA, sgRNA-A shown in purple, was designed to target the *HBA2* 3`UTR at the point where the *HBA2* and *HBA1* UTR sequences begin to diverge. The PAM sequence is underlined.
B. Plasmid map of pCas9-sgRNA-A. The human U6 promoter drives sgRNA-A expression and the Chicken beta actin short promoter and CMV enhancer (CBP) drives the expression of hSpCas9-2A-GFP. A nuclear locatisation signal (NLS) is present at the 5` and 3` end of hSpCas9.
C. T7 Endonuclease Assay. WTH3 mES cells were transfected with pCas9-sgRNA-A and a vehicle control. The primers used amplify an expected band size of 284 bp and digestion of the sgRNA-A sample with T7 Endonuclease ("+" below the gel image) gives bands of approximately 164 and 120 bp (black arrow indicates digested product). The images were analysed using ImageJ software and indel frequency was estimated to be 22%.

4.2.2.4 Mutagenesis of *HBA2* 3`UTR does not affect expression

To test the effects of the *HBA2* 3`UTR mutation included in the pAB027 HDR donor vector (Figure 4.10A), two expression vectors were generated. Each vector contained a ~2 kb *HBA2-2A-mCherry* cassette under the control of a CMV promoter (Figure 4.10B) and included (1) a wild type *HBA2* 3`UTR or (2) a *HBA2* 3`UTR with the mutated sequence used in HDR donor vector pAB027 (Figure 4.10A). The expression levels of these cassettes was then assayed by western blot and reverse transcription qPCR (RT-qPCR).

Briefly, a ~2 kb fragment containing the *HBA2-2A-mCherry* fusion gene and the entire 5`UTR and mutated 3`UTR of *HBA2* was gel purified following sequential restriction digestion of pAB027.4 with *Hpa*I and *Eag*I. The 5` overhang generated by *Eag*I digestion was blunted by treatment with Klenow enzyme. The pCDNA3.1(-) expression vector (received as a gift from Prof. Paul Digard, University of Edinburgh) was linearised by digestion with *EcoR*V and dephosphorylated to prevent re-ligation by treatment with rSAP. The two fragments were ligated overnight and 20 colonies were screened by PCR. Of the colonies screened, 5 appeared positive and were expanded and plasmid DNA was isolated. These 5 plasmid clones were screened by restriction digestion with *Sma*I. Two of these pCMV-HBA2-2A-mCherry-3`UTRmutation clones were found to have the insertion in the correct orientation and were confirmed by Sanger sequencing. A wild type pCMV-HBA2-2A-mCherry-WT3`UTR expression vector was then generated by repairing the 3`UTR sgRNA-A target site mutation in pCMV-HBA2-2A-mCherry-3`UTR mutation using site directed mutagenesis. Correct repair was confirmed by Sanger sequencing (data not shown).

The pCMV-HBA2-2A-mCherry expression vectors (Figure 4.10B) with a 3`UTR mutation and WT 3`UTR were transfected into WTH3 mES cells, alongside a vehicle control (containing only Lipofectamine reagent, OptiMEM and TE). Whole cell lysates were extracted and western blots were performed using antibodies specific to HBA and mCherry (expected band size; 15 and 27kDa, respectively), and β-actin as a loading control (expected band size; 40kDa). The HBA protein was only

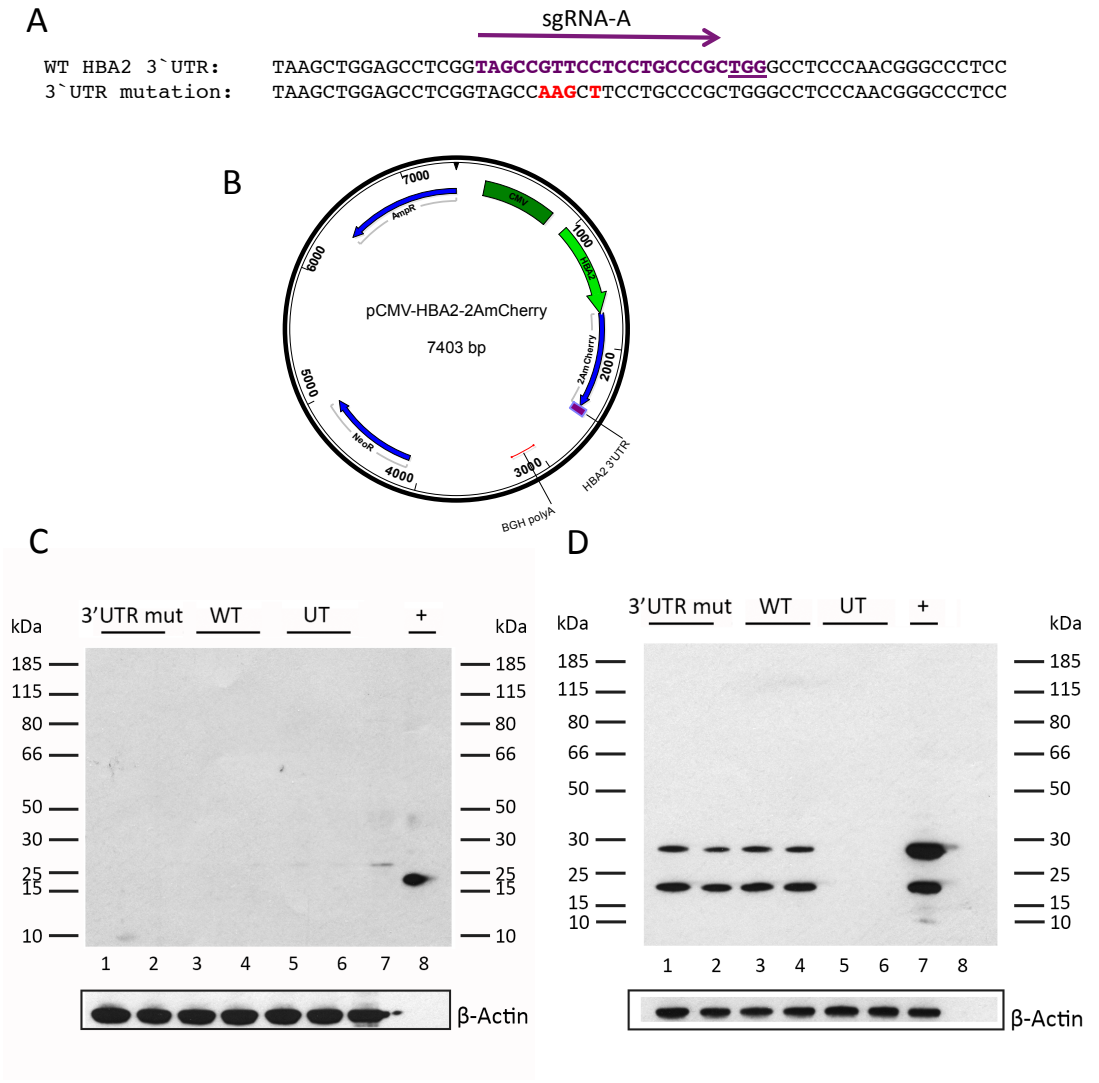


Figure 4.10 Expression validation by western blot following *HBA2* 3` UTR mutation.

A. The first 55 nucleotides of wild type *HBA2* 3`UTR following the *HBA2* stop codon are shown. The binding site for sgRNA-A is shown in purple with the PAM underlined. The 3`UTR mutation included in pAB027 is shown underneath in red.

B. Plasmid map of expression vector pCMV-HBA2-2A-mCherry.

(C-D) WTH3 cells were lipofected with constructs containing the CMV-HBA2-2A-mCherry cassette with a 3`UTR mutation (3`UTR), a WT 3`UTR (WT) or a vehicle control (UT). Western blotting was performed using antibodies for HBA **(C)** and mCherry **(D)**. HBA was not detected in any sample apart from the positive control for HBA, 0.1 µg primary erythroblasts cell lysate (lane 8) which gave the expected band size of ~15 kDa. mCherry expression was detected in all transfected cells (D) and confirmed equal expression and correct translation of mCherry, not fused to HBA2. A positive control of a lysate from HEK293 cells expressing mCherry (27 kDa, lane 7) was included. Duplicate lanes contain cell lysates from two independent biological replicates.

detected in the positive control of primary erythroblasts lysate (lane 8, Figure 4.10B). A western blot probing for mCherry was performed on the same samples and a band of approximately 27 kDa is observed in all western blot lanes containing lysates from transfected cells with expression vectors (lane 1-4, Figure 4.10C). Of note, the 27 kDa band observed in the pCMV-HBA2-2A-mCherry transfected cells is the same molecular weight as the band observed in the positive control sample (received as a gift from Prof. Paul Digard, University of Edinburgh), HEK293 cells transfected with pCMV-mCherry only (lane 7, Figure 4.10C).

WTH3 mES cells were lipofected again with pCMV-HBA2-2A-mCherry expression vectors with a 3`UTR mutation and WT 3`UTR and a vehicle control. RNA extractions were performed on mCherry positive cells isolated by FACS. RT-qPCRs (Taqman) were performed on cDNA generated from two independent biological replicates (transfections performed on different days) and one technical replicate (transfections performed on same day in different tissue culture wells). The expression levels of *mCherry* and *HBA* were normalised to the expression the housekeeping gene, *GAPDH*. The *mCherry* and *HBA* expression levels varied between the transfection replicates but, importantly, no statistically significant difference was detected between *mCherry* or *HBA* expression in the 3`UTR mutation vs. the WT 3`UTR transfected samples (Figure 4.11). As expected, *mCherry* expression was not detected in the vehicle control samples but no statistical difference was detected in *HBA* expression between the transfected samples and the vehicle control sample (Figure 4.11).

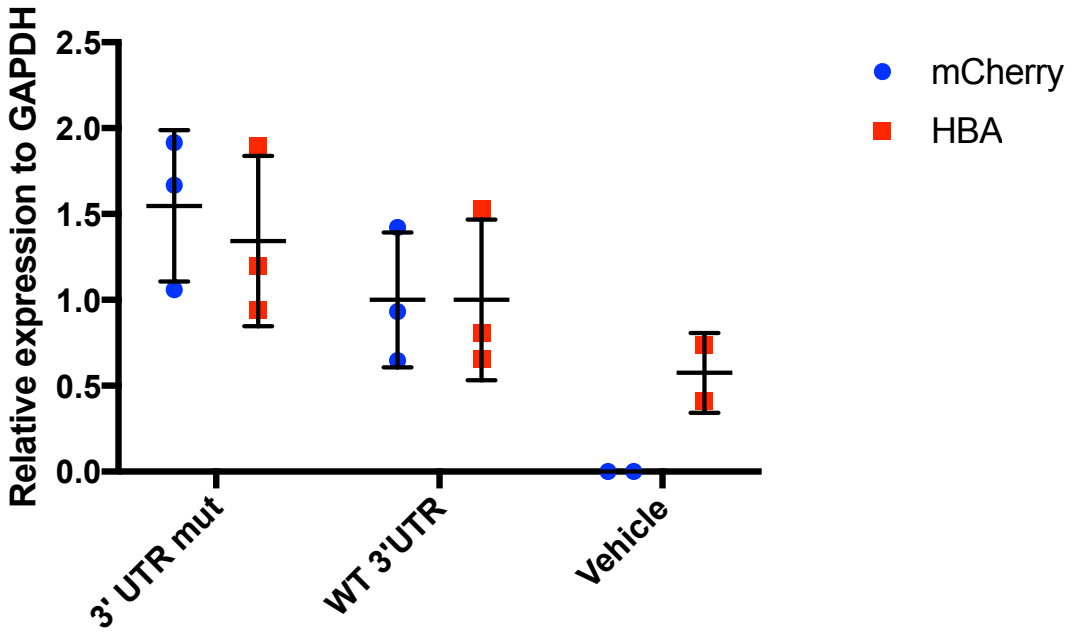


Figure 4.11 Expression validation by RT-qPCR following *HBA2* 3` UTR mutation.
WTH3 cells were transfected with constructs containing a CMV-HBA2-2A-mCherry expression cassette with a 3`UTR mutation, a WT 3`UTR or a vehicle control. RT-qPCR was performed on mCherry positive cells isolated by FACS. Paired data points from two independent biological replicates (transfections on independent days) and from one technical replicate for the 3`UTR mut and WT samples (transfection on same day in independent wells) are shown. Paired data points from two independent biological replicates for the vehicle control are also shown. A one-way ANOVA analysis showed no significant difference between *HBA* expression in the three conditions. The expression of *mCherry* and *HBA* in cells transfected with a WT 3`UTR vs. 3`UTR mutation construct cannot be distinguished by statistical analysis using a paired two-tailed t-test. The expression of *mCherry* and *HBA* is shown relative to *Gapdh* expression. The error bars represent the standard deviation from the mean**.**

4.3 Discussion

In this chapter, a number of gene targeting strategies using recombineering and CRISPR/Cas9 technologies were designed. Recombineering was successfully employed to fuse a *2A-mCherry* cassette to *HBA2* in a BAC containing the human α-globin locus. A number of CRISPR/Cas9 gene targeting tools were also generated. A CRISPR sgRNA targeting the 3`UTR of *HBA2* was designed, cloned and validated. Gibson assembly was used to generate a HDR donor vector with and without a *piggyBac* flanked counter-selective cassette. Finally, a 3`UTR mutation included in the HDR donor vectors was confirmed to preserve the expression of *HBA2-2A-mCherry*.

Recombineering/RMCE vs. CRISPR/Cas9

The recombineering steps undertaken to generate BACmCh5, although successful, were laborious and time consuming. The concentration of antibiotics used to isolate DY380 *E. coli* cells containing positively targeted BACs required extensive optimisation (data not shown) and 6 months had elapsed before the generation and confirmation of BACmCh5. A further 7 cloning steps would then have been required to generate the gene targeting and donor vectors required for RMCE in mammalian cells (Figure 4.2). A number of these cloning steps would involve additional *E. coli* recombineering steps with at least 6 gene-targeting steps in the mES cells. As this was a time-sensitive project, alternative gene targeting strategies, which had the potential to progress more rapidly, were prioritised. Nevertheless, the targeted BACmCh5 proved to be an extremely useful tool. It was used as a template for generating the CRISPR/Cas9 HDR donor vectors in this chapter and also used as a positive control for PCR screening during gene targeting in Chapter 5 (pp. 132).

Of interest, two recent publications have successfully knocked-in large genomic regions (200 kb and 42 kb, respectively) with the combined use of the Cas9 nuclease and a BAC (Baker et al., 2017; Yoshimi et al., 2016). In the first case, a ligation based method was used in which two sgRNAs, targeting either the gDNA or the BAC, two single-stranded donor oligonucleotides with homology to the gDNA,

Cas9 mRNA and a BAC with no homology to the gDNA were co-injected into rat embryos. Approximately 10% of the born pups showed correct integration of 200 kb of sequence unique to the BAC by PCR (Yoshimi et al., 2016). A more recent study from Francis Stewart's lab used lipofection to introduce (1) Cas9 expressing plasmids with sgRNAs targeting both the BAC and the gDNA and (2) an engineered BAC with long homology arms to the gDNA (4-50 kb), to mES cells (Baker et al., 2017). Although only a low number of correctly targeted clones were recovered (1-8 clones, depending on the length of homology arm used), these represented up to 14% of the antibiotic resistant clones recovered, indicating the importance of antibiotic selection cassettes in targeting screens (Baker et al., 2017). These two methods suggest another potential use for the BACmCh5 generated here, in combination with CRISPR/Cas9 technology, in future gene targeting strategies.

The use of CRISPR/Cas9 technology for gene editing became widespread from 2013 onwards following the publication of a number of high impact protocols claiming extremely accurate and fast editing efficiencies (Mali et al., 2013b; Ran et al., 2013a; Wang et al., 2013; Yang et al., 2013). Indeed, a study comparing knock-in efficiency in Zebrafish embryos using CRISPR/Cas9 or BAC technology showed that the CRISPR/Cas9 technology was 15% more efficient than the latter (Kimura et al., 2014). A number of research groups in the Roslin Institute and the University of Edinburgh have become proficient users and instructors of this technology, hence CRISPR/Cas9 technology was investigated as an alternative means of targeting the α-globin locus. The multiple CRISPR/Cas9 gene editing strategies that were designed included just 5 cloning steps to generate the HDR donor vectors and almost all of these steps involved Gibson assembly, an extremely efficient PCR based cloning technique. The number of gene targeting steps required in mES cells was between 2 and 6, depending on the strategy employed. This suggested that the CRISPR/Cas9 would provide a faster route than recombineering to successful gene editing in mES cells and so I focused on generating the tools to pursue this strategy.

Generation of CRISPR/Cas9 gene targeting tools

First, sgRNA-A was designed to specifically target the 3`UTR of *HBA2*. This guide spanned a region of the 3`UTR that has diverged between *HBA2* and *HBA1* including 7 mismatches between the two genes. It was expected that this guide would therefore be unable to anneal to, and generate indels at, *HBA1*. The guide RNA sequence was cloned into an expression plasmid containing human optimised Cas9 and a GFP reporter to generate the targeting vector, pCas9-sgRNA-A, and was validated by T7 Endonuclease I assay.

Next, two HDR donor vectors for use with the CRISPR/Cas9 strategy were generated by Gibson assembly. The first vector, pAB017, included a 5` homology arm of 800 bp and a 3` homology arm of 1300 bp to the *HBA2* locus, but did not contain a selection cassette. I also decided to generate a second HDR donor vector, pAB027, which included a 5` and 3` homology arm of 800 bp to *HBA2* and also included a *piggyBac* flanked counter-selective cassette. Although many reports have lauded the efficiency of Cas9 in producing indels (Mali et al., 2013b; Mandal et al., 2014; Wang et al., 2013), the efficiency of HDR targeting using traditional CRISPR/Cas9 techniques is quite low at ~0.5-30% (Ran et al., 2013a; Yang et al., 2013). The inclusion of a counter selective cassette could potentially increase the efficiency of gene targeting screening but also increases the number of gene targeting events required to generate a final locus with the removed selection cassette (Yusa et al., 2011).

Although CRISPR/Cas9 gene targeting can be used to knock-in sequences *via* HDR, there is a risk that the CRISPR/Cas9 complex will generate an indel mutation *via* NHEJ if the guide RNA PAM sequence or the guide seed sequence is not destroyed by integration of the donor sequence. It has been shown that the *HBA2* 3`UTR is sensitive to mutations which can easily disrupt its expression. As such, a known mutation (Weiss and Liebhaber, 1995) was incorporated at the *HBA2* 3`UTR in the HDR vector pAB027 where the sgRNA-A guide RNA was designed to anneal to. To confirm that this mutation did not negatively affect the expression of the *HBA2-2A-mCherry* transcript, two expression vectors were generated containing a *CMV-HBA2-2A-mCherry* expression cassette with a WT or mutated 3`UTR.

The WT and 3`UTR mutated expression vectors were transfected into WTH3 mES cells and western blots on whole cell lysates confirmed that mCherry was expressed at similar levels from both the wild type and 3`UTR mutation constructs (Figure 4.10D). Interestingly, a western blot using an antibody against HBA on the same cell lysate samples did not detect HBA in any of the samples of transfected cells, but HBA was detected in a positive control sample using primary erythroblasts cells (Figure 4.10C). The Alpha Haemoglobin Stabilising Protein (AHSP) is abundant in erythroid cells and has been shown to be important for stabilising higher order HBA complexes but has not been shown to be necessary for HBA monomer stabilisation *in vitro* (Kihm et al., 2002). Therefore, the absence of AHSP in mES cells should not affect the stability of HBA protein in mES cells.

The *HBA2* gene is located 5` to the *mCherry* gene in the expression construct and the two genes are separated by a 2A peptide sequence which mediates separate translation of the two mRNA sequences *via* a ribosomal skip mechanism (Szymczak et al., 2004). No 'uncleaved' HBA2-2A-mCherry protein could be detected in any of the transfected samples (expected size; 42 kDa) suggesting that 2A peptide mediated 'cleavage' efficiency is high. As the mCherry protein and transcripts are detected, it would suggest that the *HBA2* gene is also transcribed and translated but may not accumulate to detectable levels in the transfected mES cells. Although HBA mRNA or protein has been detected (Biagioli et al., 2009) or ectopically expressed (Ji et al., 2011; Weiss and Liebhaber, 1994) in non-erythroid cell lines, a literature search has not identified any reports in which HBA protein was successfully transiently expressed in mES cells. It remains unclear why transient expression of HBA could not be detected but mCherry could in the experiments described here.

RT-qPCR data showed that both *mCherry* and *HBA* transcripts were detected in transfected cells and, importantly, there was no significant difference detected by RT-qPCR between the levels of transcripts detected in the WT vs. 3`UTR mutation samples (Figure 4.11). It is also important to note, however, that no statistically significant differences were detected between the *HBA* expression in transfected and vehicle control cells. It is known that an RNA-protein complex called αCP is required to enhance *HBA* mRNA stability specifically in erythroid cells but HBA mRNA and

protein expression can still be readily detected in non-erythroid cell types that are transiently transfected with globin expression constructs (Weiss & Liebhaber 1994; Ji et al. 2007; Biagioli et al. 2009; Ji et al. 2011).

The gene targeting strategies and large number of new, validated tools generated in this chapter complete an essential step in the progress of a major goal of this thesis; genetic editing at the α-globin locus in humanised mES cell lines.

Chapter 5 Gene Targeting at the α-Globin Locus

5.1 Introduction

A number of gene targeting attempts using CRISPR/Cas9 technology targeting the α-globin locus in humanised mES cells are described in this chapter. All the targeting strategies employed attempted to fuse a *2A-mCherry* cassette to the last codon of *HBA2* in the humanised WTH3 mES cell line. The targeting vector, pCas9-sgRNA-A, was used in all targeting attempts in conjunction with either a HDR donor vector without (pAB-017) or with a selection cassette (pAB-027). A number of different transfection methods were used to introduce the constructs into mES cells and the recovered clones were screened by PCR and Southern blotting.

The optimisation of transfection efficiency is essential to ensure the availability of a large pool of transfected cells for screening during gene-targeting experiments. A number of transfection methods have been developed to introduce plasmids into mammalian cells. A traditional method used to transfect mES cells is electroporation, in which a small electric charge is passed through the cells to increase the permeability of their membranes (reviewed in Kim & Eberwine 2010). Since 2002, over 2,500 papers have been published using a modernised electroporation system called nucleofection (Lonza, http://bio.lonza.com/citations.html). Nucleofector kits have been optimised with different buffers and electroporation settings for a variety of cell types. Another transfection method, here termed lipofection, is based on the formation of cationic liposome/DNA complexes which fuse with the cell membrane and enter the cell *via* the endocytic pathway (reviewed in Maurisse et al. 2010). The varying efficiencies of these transfection techniques for use with mES cell lines has been tested and several studies show that nucleofection and lipofection techniques outperform traditional electroporation (Lakshmipathy et al., 2004; Li et al., 2014; Maurisse et al., 2010). Nucleofection, however, can result in extremely high percentages of cell death depending on the cell type used (Lakshmipathy et al., 2004).

The choice of an appropriate screening method to analyse clones isolated after transfections in gene targeting is also extremely important. Restriction fragment length polymorphism assays and PCR screening are simple methods often used in HDR screening strategies in combination with Sanger sequencing (Ran et al., 2013a; Yang et al., 2013). The more laborious Southern blotting methods developed by Edwin Southern in the University of Edinburgh (Southern, 1975) are seen as the 'gold standard' of genotyping and are still used to give a more complete view of the location and context of a genomic insertion. Other genotyping methods include qPCR and next-generation sequencing (Bell et al., 2014; Gómez-Rodríguez et al., 2008) but these are less frequently used due to their higher cost. The use of a selection cassette facilitates the efficient removal of clones lacking the appropriate antibiotic resistant gene before screening, thus increasing targeting screening efficiency (reviewed in Gonzalez 2016).

In this chapter, three transfection methods were used to introduce a CRISPR/Cas9 targeting vector and HDR donor vector, with and without a selection cassette, to WTH3 mES cells. The gene targeting attempts and screening strategies/results are summarised in Figure 5.1. In total over 650 clones were isolated, expanded and screened by a combination of PCR assays. A number of clones were taken forward for further Sanger sequencing, Southern blot and qPCR screening. Clones were identified that contained a single insertion of the *2A-mCherry* at the *HBA2* locus.

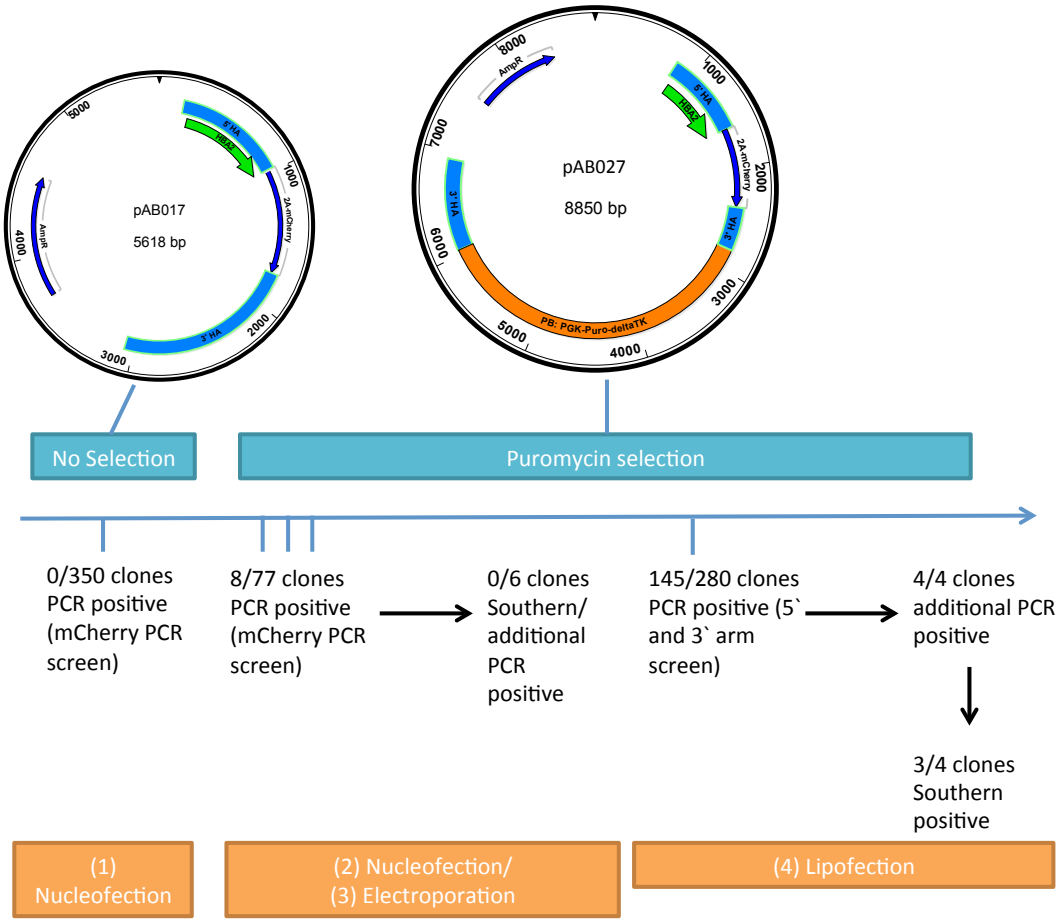


Figure 5.1 Summary of targeting strategies and results.
Timeline showing the various targeting strategies and transfection methods used with the aim of inserting the *2A-mCherry* reporter cassette into the *HBA2* locus in WTH3 humanised mES cells. Turquoise boxes indicate when positive selection was used (as determined by the HDR donor vector used). Orange boxes indicate the transfection methods used and numbers indicate which targeting strategy was used. The number of colonies screened, the outcome of screening and the screening method used are also described.

5.2 Results

5.2.1 Knock-in strategy using Cas9 and sgRNA-A in combination with pAB-017

Initially, a gene targeting strategy was designed using the validated targeting vector (see Chapter 4, pp. 122), pCas9_sgRNA-A, and the HDR donor vector pAB-017. The HDR donor vector pAB-017 included two *HBA2* homology arms and a *2A-mCherry* cassette fused to the 3` end of *HBA2* without a selection cassette. Briefly, 2 million WTH3 mES cells were co-nucleofected with 1.5 µg pCas9_sgRNA-A and 1.5 µg pAB-017.2. One day after cells were transfected, positively transfected cells (79% of total population) were isolated by FACS for GFP expression. GFP positive cells were seeded in fresh 10 cm dishes at low density (200-1000 cells/dish) and single cells were left to expand into colonies for 8 days. Approximately 350 mES cell colonies were picked into 96 well plates. Picked colonies were expanded and cells were harvested for screening by PCR.

Following gDNA extraction, screening PCRs were performed on all ~350 clones using the primers shown in orange in Figure 5.2A. These primers should give a WT band of 325 bp or a larger band of 1.1 kb following integration of *2A-mCherry*. No clones appeared to show the correct 1.1 kb integration band, however, some clones appeared to have bands of slightly greater or less than the expected 325 bp WT band (lane 10 vs. lane 11, Figure 5.2B). This suggested that an indel event occurred in these clones, which is consistent with CRISPR/Cas9 activity. PCRs were repeated on selected clones with a high fidelity Q5 polymerase and sent for Sanger sequencing. Sanger sequencing confirmed the presence of indels in the region of the sgRNA-A target site in clones 4B11 and 2B8 (Figure 5.2C).

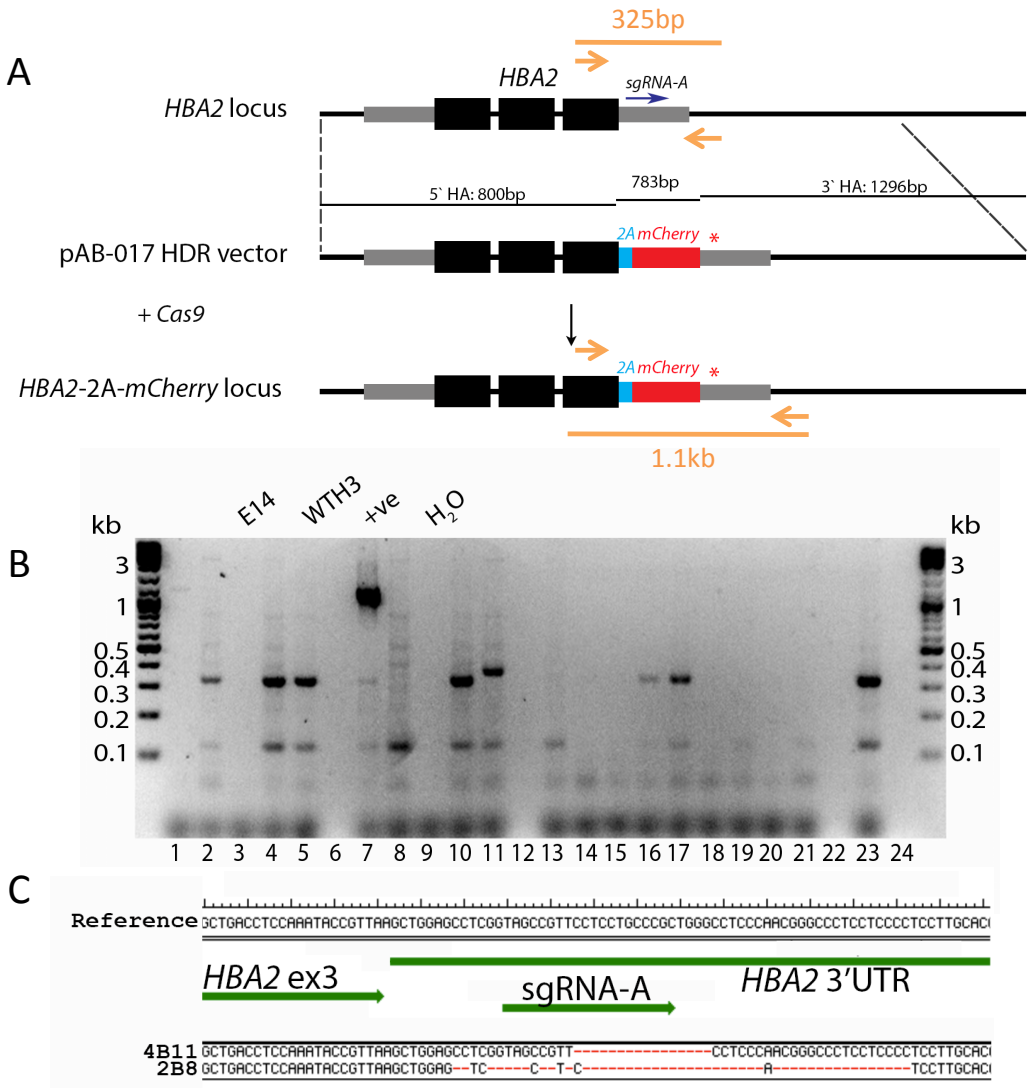


Figure 5.2 Knock-in strategy using sgRNA-A in combination with pAB-017.
A. Schematic showing knock-in strategy using HDR donor vector pAB-017, a vector without a selection cassette, in combination with sgRNA-A (blue arrow) and Cas9. The orange arrows show a primer pair used in PCR screening in B. These were designed to generate a WT band at 325 bp and a band indicating a positive insertion at 1.1 kb. The occurrence of both bands would suggest a random integration of the donor vector has occurred.
B. Representative gel image from PCR screening. Approximately 350 clones were screened and no integration events were detected. Certain clones gave slightly larger (lane 11) or smaller bands (not shown) compared to the WT band (lane 5). The gDNA control WTH3 gave a WT band of 325bp (lane 5). As a positive control for the 1.1 kb band, pAB17.2 was mixed with WTH3 gDNA at equimolar concentrations (lane 7). Negative controls gDNA from E14 cells and H_2O (lane 3 and 9, respectively) only gave no apparent bands. All lanes that are unlabeled above the gel contain individual clones in screening. *Figure legend is continued on the next page.*

5.2.2 Knock-in strategy using Cas9 and sgRNA-A in combination with pAB-027

5.2.2.1 Targeting strategy using pAB-027

As co-transfection of a HDR vector without a selection cassette (pAB-017) and pCas9_NucA did not yield clones with an integration of the *2A-mCherry* cassette, the targeting was repeated with a new HDR vector containing a selection cassette (pAB-027) in order to improve screening efficiency (Figure 5.3). The selection cassette included in pAB-027 is a *piggyBac* inverted repeat flanked PGK driven *puromycin-Δthymidine kinase* cassette (*PB:PGK-PuroΔTK*). This cassette allows for positive selection of clones with correct integration using puromycin selection. Following transient *piggyBac* transposase transfection clones with the selection cassette seamlessly removed can be isolated by negative selection with fialuridine (FIAU). A large array of PCR screening primer sets and probes for Southern blot screening were developed to ensure correct integration of the *2A-mCherry* and selection cassette at the *HBA2* locus.

Figure 5.2 *Legend continued:*
C. PCR products from two clones with unexpected PCR product sizes (4B11 and 2B8-not shown in gel) were sent for sequencing. Results indicated that small deletions of variable sizes occurred surrounding the predicted sgRNA-A binding site.

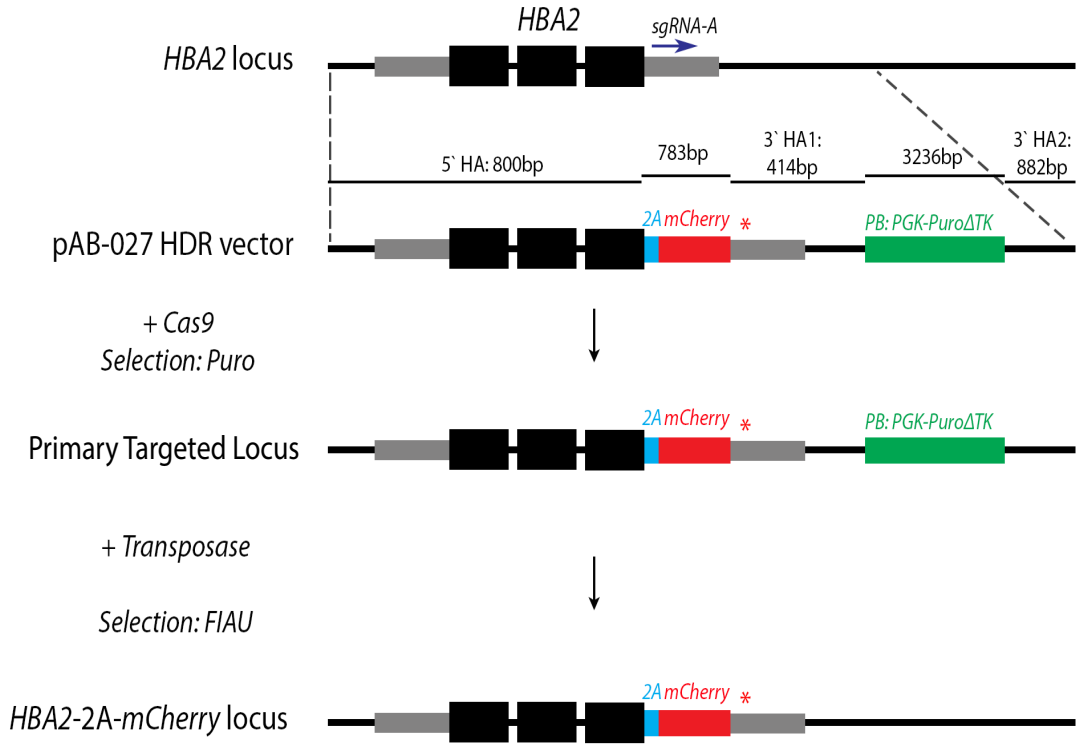


Figure 5.3 Knock-in strategy using sgRNA-A in combination with pAB-027. Targeting strategy using HDR vector pAB-027 in combination with sg-RNA-A (blue arrow). The plasmid pAB-027 is a HDR donor vector including the *PB:PGK-PuroΔTK* cassette and a 5` and two 3` homology arms (HA). Transfected cells are selected by resistance to puromycin and screened by PCR and Southern blot to ensure correct integration of the *2A-mCherry* and selection cassette. Correctly targeted cells are transfected with a *PiggyBac* transposase. Cells are selected by resistance to FIAU and sensitivity to puromycin and screened by PCR and Southern blot to ensure correct integration of the 2A-mCherry and selection cassette.

5.2.2.2 Screening of potential *2A-mCherry* knock-in clones following nucleofection/electroporation using HDR vector pAB-027

Briefly, 10 nucleofections of 2 million WTH3 mES cells with 1.5 μg pCas9_sgRNA-A and 1.5 μg pAB-027 were performed. One day later, cells were seeded in fresh 10 cm dishes initially at low density (200-1000 cells/dish), but later experiments at high density (1,000,000 cells/dish) when it became clear that the targeting efficiency (and thus the numbers of puromycin resistant clones) was low. Cells were plated in mES cell media supplemented with 1 μg/ml puromycin and single cells were left to expand into mES cell colonies for 8-14 days. A total of 16 puromycin resistant mES cell colonies were recovered into 96 well plates and expanded for freezing down cell stocks and for PCR genotyping.

In order to recover a higher number of PCR positive puromycin resisitant clones, two electroporations were performed. Briefly, 10 million WTH3 mES cells were co-electroporated with 5 μg pCas9_sgRNA-A and 5 μg pAB-027. One day later cells were seeded in fresh 10 cm dishes (1 million cells/dish) in mES cell media supplemented with 1 μg/ml puromycin and single cells were left to expand into mES cell colonies for 8-14 days. A total of 61 puromycin resistant mES cell colonies were recovered into 96 well plates and expanded for freezing down cell stocks and for PCR genotyping.

Following gDNA extraction, screening PCRs were performed on all 77 puromycin resistant clones recovered from nucleofections and electroporations using primer set 1 (mCherry PCR) shown in orange in Figure 5.4A. These primers were expected to give a WT band of 325 bp or a larger band of ~1.1 kb following integration of *2A-mCherry*. Of the 77 puromycin resistant clones tested, 8 clones were positive for a band at ~1.1 kb, indicating integration of *2A-mCherry* (lane 2, Figure 5.4Bi). Of these, 6 clones survived cell expansion and were brought forward for further screening (clones NA5 and NB1 from nucleofections and clones EA1, EB7, EB12 and ED9 from electroporations). Clone EA2 from electroporations was puromycin resistant but PCR negative and was used as a negative control in further screening.

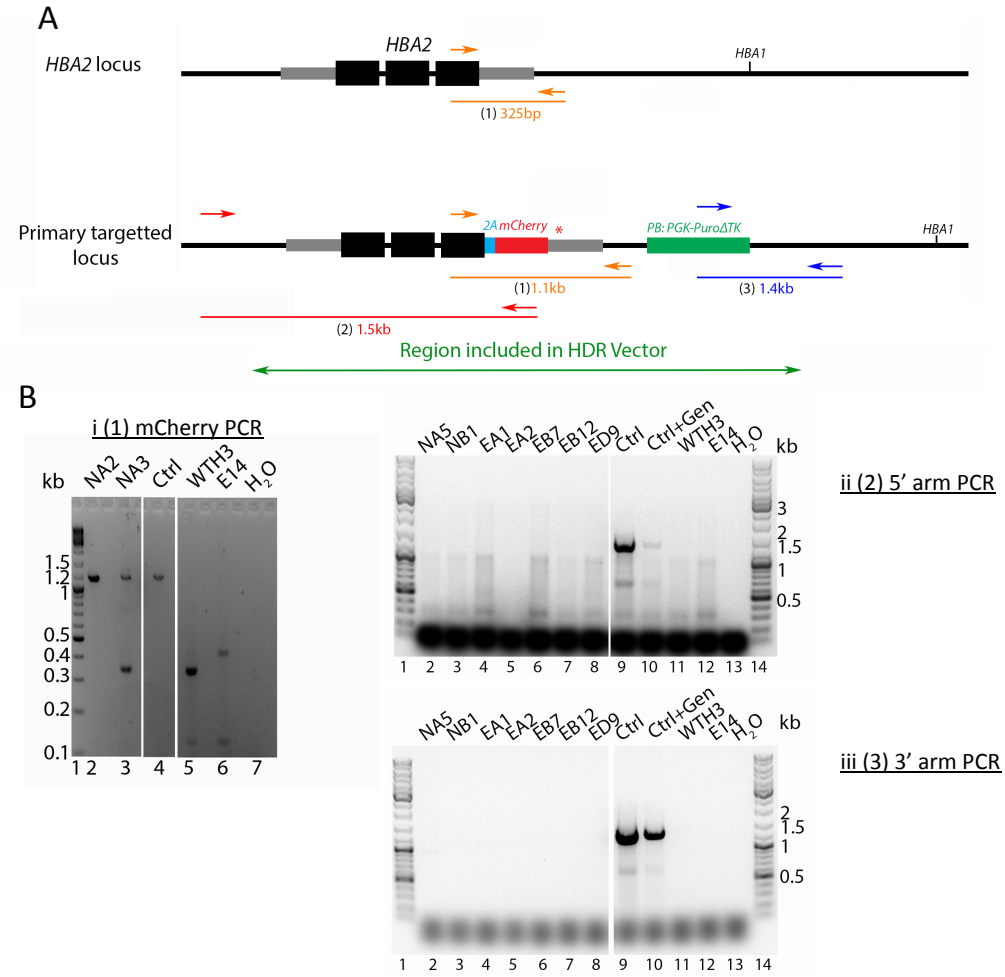


Figure 5.4 PCR screening of potential *2A-mCherry* knock-in clones following nucleofection/electroporation with pCas9-sgRNA-A and pAB-027.

A. PCR screening strategy for testing the puromycin resistant clones. Primer set 1 (orange arrows) is the only primer set that amplifies a band in the wild type *HBA2* locus. Primer set 2 and 3 (red and blue arrows) are designed to specifically amplify bands in the primary targeted locus.

B. Composite gel images showing examples of PCR screening results from puromycin resistant clones. **(i)** *mCherry* insertion PCR screening was performed using primer pair 1. Clone NA2 appears to contain an insertion whereas NA3 appears to have both a wild type and insertion band. Clone NA2 did not survive cell expansion and so was not tested in future screens. BACmCh5 was used as a control for insertion (1.1 kb band, lane 4) and WTH3 gDNA was used as a control for non-insertion (325 bp band, lane 5). Negative controls E14 gDNA (lane 6) and H_2O (lane 7) did not yield a PCR product or produced a background signal. **(ii)** A 5` arm PCR screen was performed using primer pair 2. Positive control DNA (BACmCh5 only; lane 9) mixed at equimolar concentrations to 100 ng WTH3 gDNA (lane 10) gave expected insertion band size of 1.5 kb. Negative controls WTH3 and E14 gDNA and H_2O (lanes 11-13) and all clones screened (lanes 2-8) did not yield a PCR product or produced a background signal. *Figure legend is continued on the next page.*

A 5`arm screening PCR was designed using primer set 2 (red arrows) indicated in Figure 5.4. This screen was designed to amplify a PCR product of ~1.5 kb following site-specific correct integration of *2A-mCherry* using a forward primer that was located outwith the region included in the HDR vector and a reverse primer specific to the *2A-mCherry* insertion. These primers were optimized using a template of BACmCh5 DNA mixed at equimolar concentration to 100 ng of WTH3 gDNA (lane 12, Figure 5.4Bii). A 3`arm screening PCR was designed using primer set 3 (blue arrows) in Figure 5.4A. This screen was designed to amplify a PCR product of ~1.4 kb following site specific correct integration of *PB:PGK-puroΔtk* using a forward primer that was specific to the selection cassette and a reverse primer located outwith the region included in the HDR vector. These primers were optimised on a GeneBlock (IDT) containing primer target sites mixed at equimolar concentration to 100 ng of WTH3 gDNA (lane 12, Figure 5.4Biii). Both the 5` and 3` arm PCR screens indicated that none of the 6 mCherry PCR positive clones contained the correct insertion (Figure 5.4Bii and iii).

DNA probes for Southern blotting were designed to hybridise to a region within the *HBA2-2A-mCherry* locus either specific to the *2A-mCherry* inserted cassette (mCherry probe) or beyond the homology arms included in the HDR vectors used in this thesis (5` probe, Figure 5.5A). The probes were designed such that Southern blots would reveal the specific location of the *2A-mCherry* cassette as well as the integrity of the surrounding *HBA2* locus following homology directed repair. The probes were generated by either restriction digestion of pUC_ex3HBA2_2A-mCherry-3`UTR (mCherry probe) or by PCR from a DNA template of BACmch5 (5`probe). Due to the repetitive nature of the alpha globin locus, the 284 bp 5` probe had the potential to bind both a site 3.3 kb 5` of *HBA2* (5` Probe, Figure 5.5A) and a

---

Figure 5.4 *Legend continued:*
B iii. A 3` arm PCR screen was performed using primer pair 3. Positive control DNA (lane 9; synthesised 1.6 kb GeneBlock from IDT containing sequence from puromycin cassette and extending beyond the 3` homology arm) mixed at equimolar concentrations to 100 ng WTH3 gDNA (lane 10) yielded a PCR product consistent with the expected insertion band size of 1.4 kb. Negative controls WTH3 and E14 gDNA and H_2O (lanes 11-13) did not yield a PCR product. All clones screened (lanes 2-8) appeared negative.

second site 0.7 kb 3` of *HBA2* (5` Probe*, Figure 5.5A) which is 98% homologous to the first site. Southern blots were performed on the mCherry PCR positive clones using the mCherry and 5` probes. A single band of approximately 9 kb was present when gDNA from clones EB12 and ED9 was tested with the mCherry probe (Figure 5.5B). Two bands of approximately 9 kb and 5.5 kb were present on the Southern blot with the 5` probe of DNA from clone NB1, EA1 and ED9. A single wild type band of ~14 kb was present on the Southern blot with the 5` probe of DNA from clone EB12 (Figure 5.5C). According to the PCR and Southern blot data, more clones screened from the nucleofections and electroporations contained the correct integration of the *2A-mCherry* cassette. These results are summarised in Table 5.1.

Table 5.1 Summary of screening results of puromycin resistant clones recovered after nucleofection/electroporation. The results from each PCR and Southern blot screen are listed as "Yes" for a positive result and "No" for a negative result (no band detected or incorrect banding pattern detected). A total of 8 puromycin resistant clones were recovered that were positive for PCR 1 and 6 of these survived cell expansion and were screened further. No clone was positive for all screening methods.

Transfection Method	Nucleofection				Electroporation				
Clone ID	NA2	NA4	NA5	NB1	EA1	EA2	EB7	EB12	ED9
PCR (1) mCherry	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
PCR (2) 5` arm	-	-	No	No	No	No	No	No	No
PCR (3) 3` arm	-	-	No	No	No	No	No	No	No
Southern mCherry Probe	-	-	No	No	No	No	No	Yes	Yes
Southern 5` Probe	-	-	No	No	No	No	No	No	No

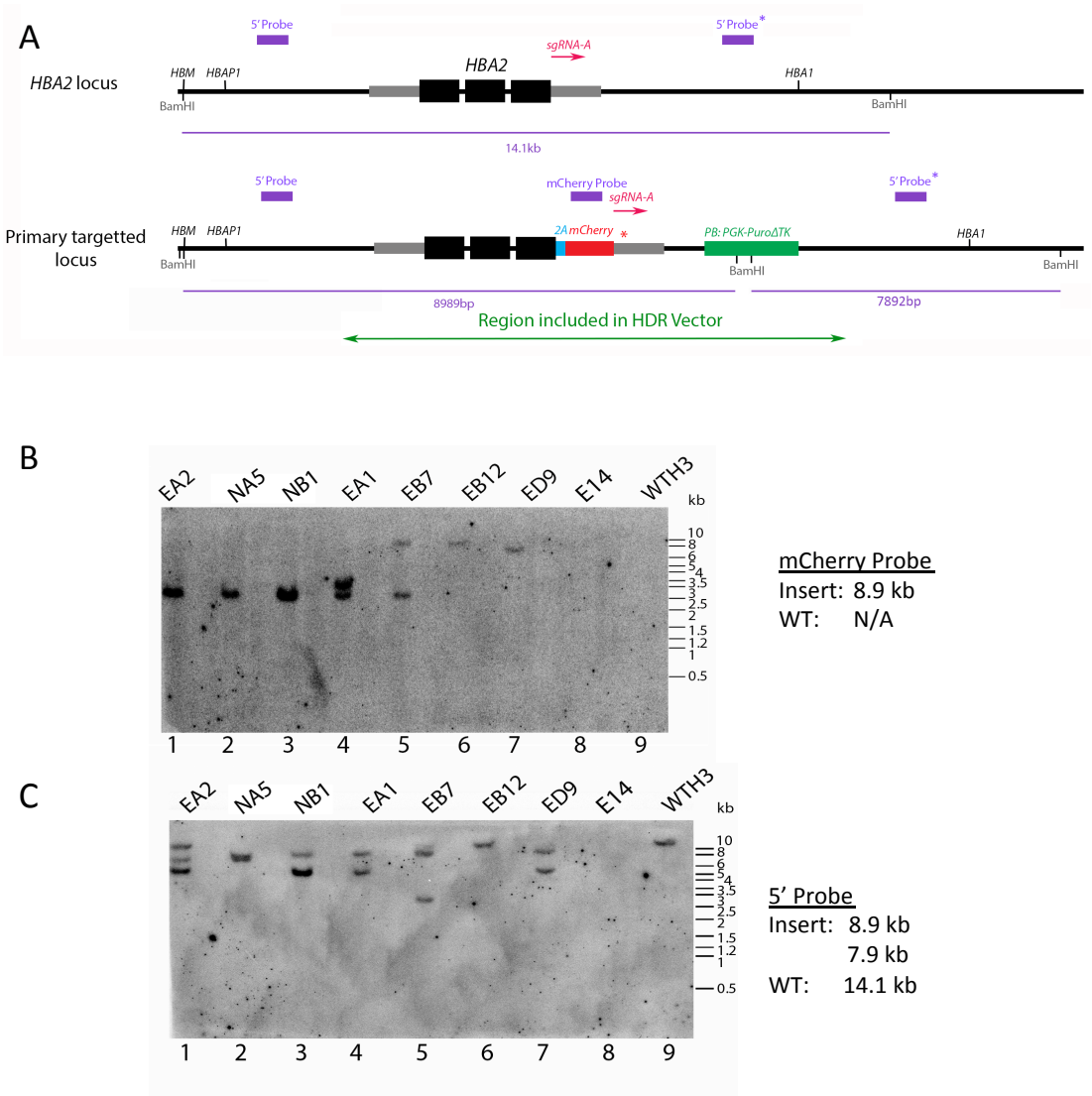


Figure 5.5 Southern blot screening of potential *2A-mCherry* knock-in clones following nucleofection/electroporation with pCas9-sgRNA-A and pAB-027.
A. A schematic of the primary targeted locus indicating *Bam*HI restriction sites and the locations of two Southern blot probes (purple boxes).
B. Results of Southern blot using the mCherry probe. No signal was detected for negative controls E14 and WTH3, as expected. Correct knock-in clones are expected to have one band at 8.9 kb. EB12 and ED9 have approximately the correct band size and EB7 appears as a doublet. A second mCherry Southern blot was performed and gave similar results (data not shown).
C. Results of Southern blot using 5` probe. WTH3 showed correct expected band size of 14.1 kb. No signal was detected for negative control E14, as expected. Correct knock-in clones are expected to have two bands at 8.9 and 7.9 kb, due to *Bam*HI restriction sites in the *Puro* gene. None of the DNA screened resulted in bands consistent with appropriate integration of the *2A-mCherry* cassette.

5.2.2.3 PCR screening of potential *2A-mCherry* knock-in clones following lipofection

As nucleofection and electroporation of WTH3 cells with pAB-027 and pCas9_sgRNA-A continued to yield very low numbers of puromycin resistant clones, which did not contain the correct insertion, lipofection was tested as a method to transfect the cells. Briefly, 5.6 million WTH3 mES cells were lipofected in suspension with 15 μg pCas9_sgRNA-A and 15 μg pAB-027.4. Cells were harvested and seeded in fresh 10 cm dishes initially at high density (1 million cells/dish) 24 h after transfection. Harvested cells were plated in mES cell media supplemented with 1 μg/ml puromycin and single cells were expanded into mES cell colonies for 9 days. A total of ~280 puromycin resistant mES cell colonies were recovered into 96 well plates and expanded for freezing down cell stocks and for PCR genotyping.

The mCherry PCR using primer set 1 in Figure 5.6. A previously identified clones which were found to be negative after further interrogation (Table 5.1). Therefore it was decided to perform initial screening PCRs using primer set 2 (5` arm) and primer set 3 (3` arm) shown in Figure 5.6A. Following gDNA extraction, initial screening PCRs were performed on all ~280 puromycin resistant clones recovered from lipofections. The 5` and 3` screening primers were expected to yield bands of 1.5 and 1.4 kb, respectively, following integration of *2A-mCherry*, without the presence of a WT band. The expected band size was amplified from 158 clones with the 5` arm PCR, 223 clones with the 3` arm PCR and 145 of these clones were positive for both 5` and 3` PCRs (double positive). Four double positive clones (L1B7, L1C5, L1C6, L1C10), one 5` only positive clone (L1B2), one 3` only positive clone (L1A10) and one puromycin resistant clone, which was negative for both PCR screens (L1A11) (Figure 5.6Bii and iii), were expanded and interrogated by further screening. Given the previous difficulty in isolating a positive clone, two additional PCR screens were developed. The four double positive PCR clones were positive for two PCR screens spanning the *2A-mCherry* cassette and the region between *mCherry* and *PB:PGK-PuroΔTK* cassette (Figure 5.6Biv and v). A final PCR screen spanning the *PB:PGK-PuroΔTK* cassette was designed to amplify a wild type region of 1.9 kb and 5.1 kb across the insertion (Figure 5.6A). Although the positive control insert band was not detected (possibly due to difficultly of PCR

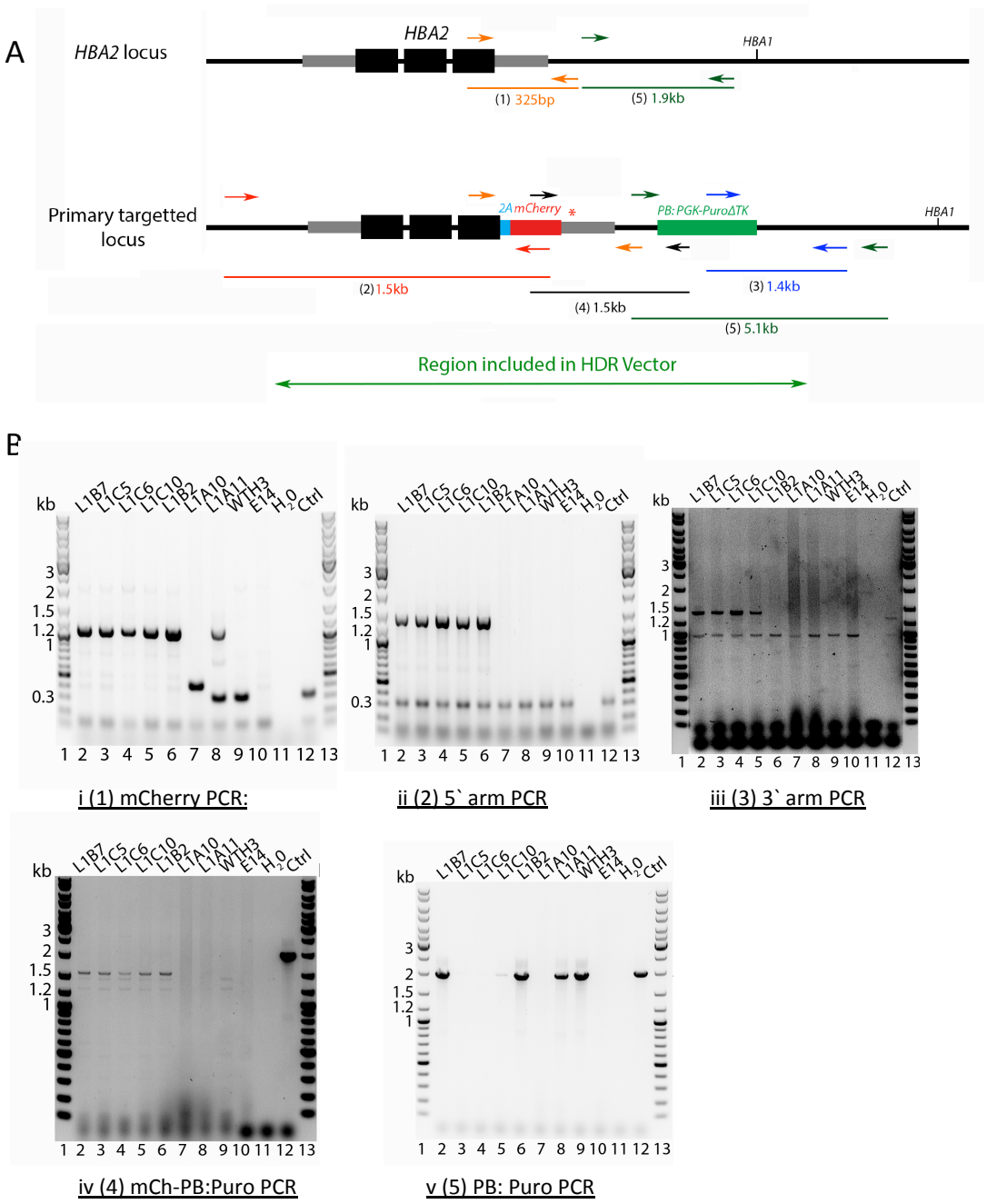


Figure 5.6: PCR screening of potential *2A-mCherry* knock-in clones following lipofection of WTH3 with pCas9-sgRNA-A and pAB-027.

A. PCR screening strategy for the Puromycin resistant clones. Primer set 1 and 5 (orange and green arrows) are the only primer sets that amplify a band in the wild type *HBA2* locus. Primer set 2, 3 and 4 (red, blue and black arrows) should only amplify bands in the primary targeted locus.

B. Gel images showing PCR screening results using primer sets **(i)** mCherry PCR **(ii)** 5` arm PCR; **(iii)** 3` arm PCR; **(iv)** mCherry-PB Puro PCR and **(v)** PB-Puro PCR. Clones L1B7, L1C5, L1C6 and L1C10 are positive for primer screens (i-iv). Of these, only clones L1C5 and L1C6 do not show amplification of WT PCR product in screen (v). *Figure legend is continued on the next page.*

amplification across *piggybac* repeats), a wild type band was detected in clones L1B7 and L1C10 suggesting that these could be mixed clones (Figure 5.6Bv). The PCR screening data (summarised in Table 5.2) suggested that clones L1B7, L1C5, L1C6 and L1C10 contain a *2A-mCherry* and selection cassette at the correct position in the *HBA2* locus and that L1C5 and L1C6 are pure clonal populations.

5.2.2.4 Southern blot screening of candidate *2A-mCherry* knock-in clones following lipofection

The candidate knock-in clones L1B7, L1C5, L1C6, L1C10, and three negative clones, including the 5` arm only positive clone (L1B2), 3` arm only positive clone (L1A10) and one puromycin resistant clone that was negative for both PCR screens (L1A11), were screened by Southern blot using the mCherry and 5` probes (Figure 5.7A). A single band at approximately 9 kb was detected in all four potential knock-in clones with the mCherry probe (Figure 5.7B) suggesting the *2A-mCherry* and selection cassette integrated at the correct genomic locus. Using the 5` probe doublets of the correct expected size were detected in clones L1C5, L1C6 and L1C10 (Figure 5.7C). A triplet was detected in clone L1B7 which indicated a large scale genomic rearrangement had occurred. A single WT band at approximately 14 kb was detected in the WTH3 sample using the 5` probe and no band was detected in the E14 negative control sample, which suggested that the probe was annealing to the correct genomic locus. The detection of correctly sized doublets in the potential knock-in clones suggested that a correct integration of the *2A-mCherry* and selection cassette had occurred. The mCherry and 5` probe Southern blots were performed two and three times, respectively, with each blot showing the same banding pattern. The results of PCR and Southern blot screening are summarised in Table 5.2.

Figure 5.6B *Legend continued:*
The contrast was increased in **B (iii)** in lanes 2-12. A positive control was run for each PCR in lane 12. The positive control DNA consisted of 100 ng WTH3 gDNA with equimolar concentrations of **(i, iv, v)** pAB-027, **(ii)** BACmCh5 and **(iii)** a synthesised 1.6 kb GeneBlock from IDT containing sequence from the *PB:PGK-PuroΔTK* cassette to outside the 3` homology arm. The positive control failed in screen **(i)** and **(ii)** but PCR products of the expected size were successfully amplified using this screen.

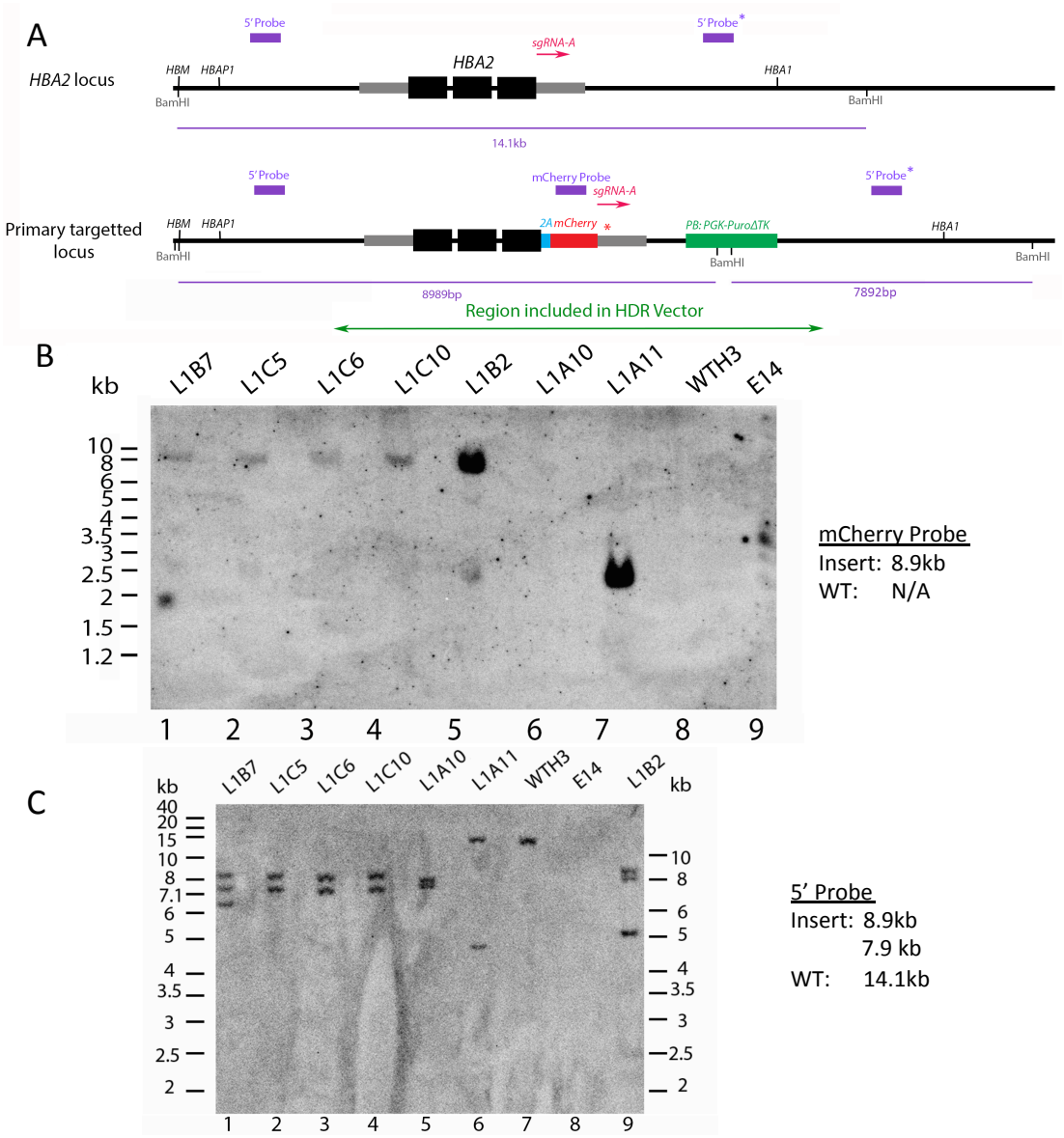


Figure 5.7 Southern blot screening of potential *2A-mCherry* knock-in clones following lipofection of WTH3 with pCas9-sgRNA-A and pAB-027.
A. A schematic of the wild type and primary targeted *HBA2* locus showing *BamH*I restriction sites and the locations of two Southern probes (purple boxes).
B. Results of Southern blot using mCherry probe. WTH3 and E14 negative controls showed no signal, as expected. A single band at the expected size was observed in clones L1B7, L1C5, L1C6, L1C10 and L1B2 suggesting the *mCherry-PB:PGK-PuroΔTK* cassette was inserted once in the correct genomic position. The negative clone L1A11 contains a single band at an incorrect size suggesting a single insertion occurred at an incorrect genomic location. A second mCherry Southern blot was performed and gave similar results (data not shown). *Figure legend is continued on the next page.*

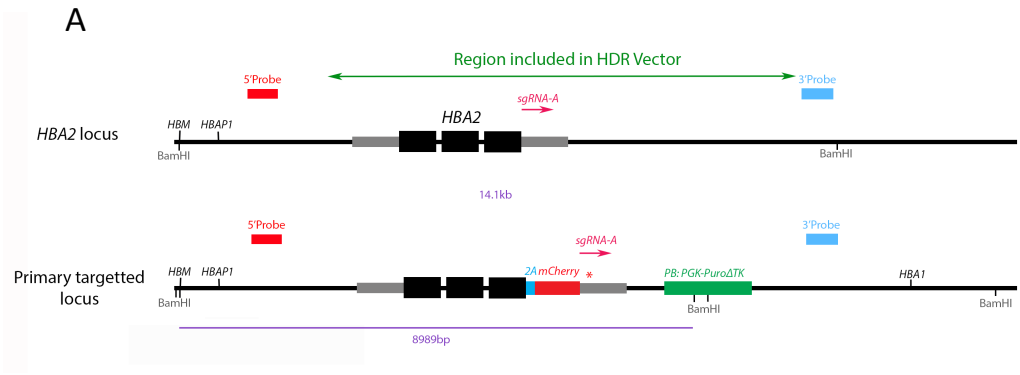
Table 5.2 Summary of screening results of puromycin resistant clones recovered after lipofection. The results from each PCR and Southern blot screen are listed as "Yes" for a positive result and "No" for a negative result (no band detected or incorrect banding pattern detected). A total of 7 puromycin resistant clones were screened. Clones L1C5 and L1C6 were positive for all screening methods.

Transfection Method	Lipofection						
Clone ID	L1B7	L1C5	L1C6	L1C10	L1B2	L1A10	L1A11
PCR (1) mCherry	Yes	Yes	Yes	Yes	Yes	Yes	No
PCR (2) 5` arm	Yes	Yes	Yes	Yes	Yes	No	No
PCR (3) 3` arm	Yes	Yes	Yes	Yes	No	No	No
PCR (4) mCherry-PB:Puro	Yes	Yes	Yes	Yes	Yes	No	No
PCR (5) PB: Puro	No			No	No		No
Southern mCherry Probe	Yes	Yes	Yes	Yes	Yes	No	No
Southern 5` Probe	No	Yes	Yes	Yes	No	No	No

Figure 5.7 *Legend continued:*
C. Results of Southern blot using the 5` probe. The WTH3 positive control gave the correct band size of ~14 kb indicating the probe successfully annealed to the target with no background signal. E14 negative control gave no signal. A triplet was observed in clone L1B7 and negative clone L1B2. Two bands of the correct expected sizes (8.9 and 7.9 kb) were observed for clones L1C5, L1C6 and L1C10. Incorrect sized doublets were observed in the negative clones L1A10 and L1A11. This data suggests that the correct insertions had occurred in clones L1C5, L1C6 and L1C10. Two further 5` probe Southern blots were performed and gave similar results (data not shown).

5.2.2.5 A qPCR screening strategy of candidate *2A-mCherry* knock-in clones

A qPCR screening strategy was developed to confirm no duplications had occurred in the α-globin locus following correct integration of the *2A-mCherry* cassette. Taqman primer sets and probes were designed to target the region included in the Southern blot 5` probe, a region downstream of that included in the HDR vectors (3` probe) and at the HS-40 enhancer ~60 kb upstream of the *HBA2* gene (Figure 5.8A). No increase was detectable by qPCR in the copy number variation of the 5` probe region in the clones L1C5, L1C6 and L1C10 (Figure 5.8B). As expected, the only clone that potentially showed an increase in copy number by qPCR at the HS-40, 5` and 3` probe region was L1B7 (Figure 5.8B), the clone in which a triplet was detected by Southern blot using the 5` probe.



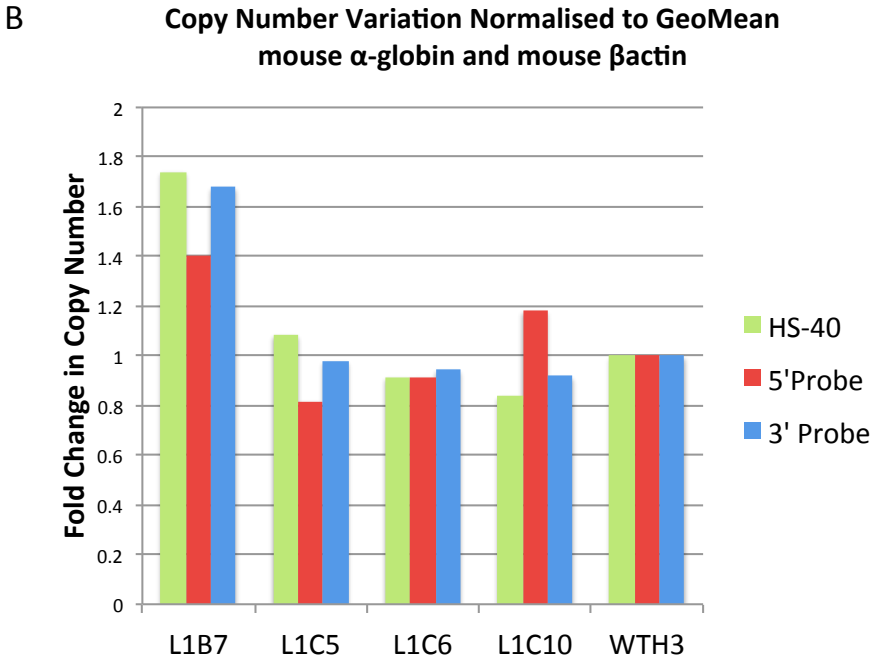


Figure 5.8 QPCR Screening of candidate knock-in clones following lipofection of WTH3 mES cells with pCas9-sgRNA-A and pAB-027.
A. Map showing the locations of Taqman qPCR probes to assess copy number variation of 5` and 3` probe regions. The HS-40 probe (not shown) is located ~60 kb upstream of the HBA2 gene.
B. Results from qPCR screening of copy number variation in the HS-40, 5` probe and 3`probe regions, normalised to the geometric mean of mouse α-globin (ex2) and mouse β-actin (promoter) in WTH3 clones.

5.2.2.6 Phenotypic analysis of candidate *2A-mCherry* knock-in clones following lipofection

The potential knock-in clones L1B7, L1C5, L1C6, L1C10 did not express detectable levels of mCherry at the mES cell stage, as expected (data not shown). *In vitro* differentiation of these clones was performed using the combined method (Chapter 3, pp. 94) and all clones were found to express levels of mCherry detectable by fluorescent microscopy by Day 7 of *in vitro* differentiation (Figure 5.9). *In vitro* differentiated EBs were harvested and mCherry levels were quantified by flow cytometry. Three clones showed mCherry expression in over 10% of the EB cell population (L1B7, L1C5, L1C10) while L1C6 showed mCherry expression in only 2.4% of EB cells (Table 5.3). The WTH3 mES cell line showed no expression of mCherry after *in vitro* differentiation (Table 5.3, Figure 5.9).

Table 5.3 Results of flow cytometry using candidate knock-in clones. Clones were differentiated in vitro towards the erythroid lineage using the combined method and flow cytometry was performed on harvested EBs. All knock-in clones showed mCherry expression following differentiation whereas the negative control parental clone, WTH3, did not express mCherry.

Cell Line	% mCherry Positive
L1B7	12.1
L1C5	11
L1C6	2.4
L1C10	16.7
WTH3	0

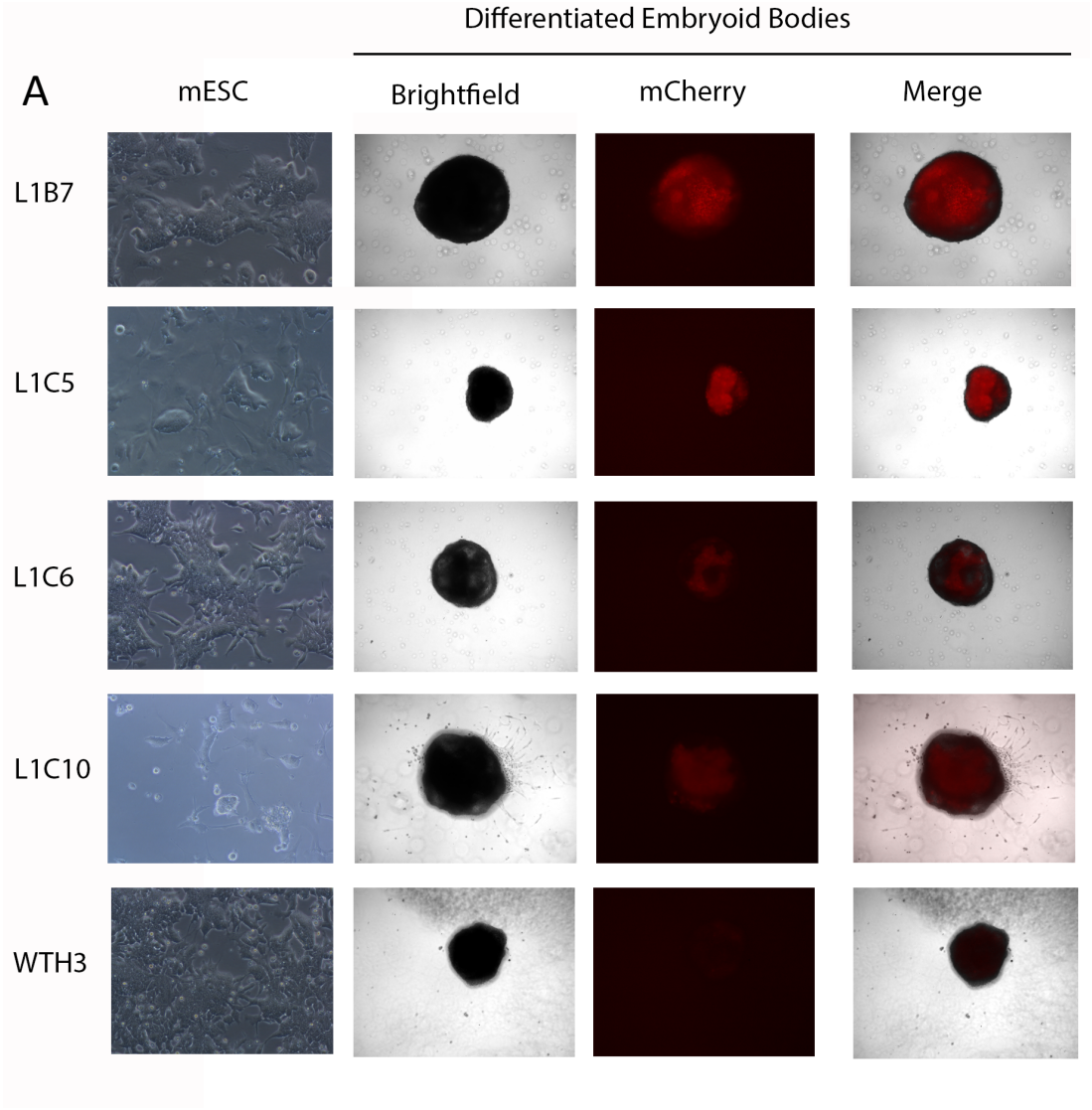


Figure 5.9 Phenotypic analysis of candidate knock-in clones by *in vitro* differentiation and fluorescent microscopy. Clones were differentiated *in vitro* towards the erythroid lineage using the combined method and imaged on Day 7 of differentiation. All clones expressed mCherry following differentiation unlike the negative control parental clone, WTH3. The mES cell images (mESC) were taken using a Zeiss axiovert 25s inverted microscope at 20x magnification. The EB images were taken using a Zeiss axiovert 100 inverted microscope at 10x magnification. The fluorescent mCherry EB images were taken using a TRITC/dsRED filter and merged using the ImageJ software package.

5.3 Discussion

In this chapter, I made multiple attempts to fuse a *2A-mCherry* cassette to the *HBA2* gene in humanised mES cells using CRISPR/Cas9 technology. A guide RNA, sgRNA-A was used in combination with two HDR donor targeting vectors, with and without a selection cassette, and three different methods of transfection were tested. An extensive array of screening methods were employed to analyse the recovered mES cell clones and of the approximately 650 clones analysed, three clones (L1C5, L1C6 and L1C10) were recovered that had an insertion event without the occurrence of apparent chromosomal rearrangement events within the locus. The methods used and the results obtained for gene targeting are summarised in Table 5.4.

Table 5.4 Summary of gene targeting methods and results. Three methods of transfection were used in gene targeting; nucleofection (nuc), electroporation (elec) and lipofection (lipo). Positively transfected cells were isolated by FACS for strategy 1 and by puromycin resistance (Puro) for strategy 2-4. For strategy 1-3, PCR screen 1 was first used to isolate positive clones. For strategy 4, PCR screen 2 and 3 were first used to identify 145 clones positive for both PCR screens. Out of 7 of these double PCR positive clones, 5 were positive for PCR screen 1. Three clones were positive for both Southern blot screening strategies.

Strategy	HDR Donor Vector	Transfected Cell Number	Method	Selection	Number of clones isolated	Positive clones (PCR screen)			Positive clones Southern blot
						1. mCherry	2. 5' arm	3. 3' arm	
1	pAB017	2 x 10^6	Nuc	FACS	~350	0	-	-	-
2	pAB027	20 x 10^6	Nuc	Puro	16	3	0	0	0
3	pAB027	20 x 10^6	Elec	Puro	61	5	0	0	0
4	pAB027	2.6 x 10^6	Lipo	Puro	~280	(5)	158	223	3

Gene targeting and screening

The first gene targeting strategy used the HDR donor vector without a selection cassette, pAB-017, in combination with sgRNA-A. Of the 350 clones recovered, no clones were identified that had a correct insertion event, as assessed by PCR screen 1. However, a number of clones were isolated with indels at the *HBA2* 3`UTR, suggesting that sgRNA-A was successfully guiding Cas9 to its target to cause a double stranded break, which was then repaired by the error-prone NHEJ

pathway. This confirmed that sgRNA-A was appropriate to use in targeting the *HBA2* 3`UTR, however a new targeting strategy to increase the efficiency of selection following HDR was required.

The second targeting strategy I employed used nucleofection of pCas9-sgRNA-A in combination with a HDR donor vector that contained a piggyBac flanked counter selective cassette, pAB-027. The incorporation of a selection cassette was expected to increase selection efficiency and decrease the number of clones required for screening (Gonzalez, 2016; Yusa et al., 2011). Another apparent benefit of the incorporation of a selection cassette was that positively transfected cells did not require sorting by FACS, a costly and time consuming technique. Instead, transfected cells could be transferred after 24 h to media supplemented with 1 μg/ml puromycin to directly select for the integration of the selection cassette. Unfortunately, extremely low numbers of puromycin resistant clones were recovered using this strategy on multiple occasions (16 clones isolated from 10 nucleofections transfecting 20 x 10^6 mES cells). It has been reported that adult stem cells also have extremely high cell death rates following nucleofection but that the use of electroporation results in higher rates of cell recovery (Lakshmipathy et al., 2004). Therefore, electroporation was tested as a means of transfection (strategy 3, Table 5.4). Indeed, a higher number of puromycin resistant clones were recovered when cells were electroporated vs. nucleofected (61 clones from 2 electroporations transfecting 20 x 10^6 mES cells). Although 8/77 puromycin resistant clones isolated using strategy 2 and 3 were positive for PCR screen 1 using primers flanking the *2A-mCherry* insertion site, none of these clones were positive for additional PCR screens or Southern blot screens.

Finally, a fourth gene targeting strategy was implemented and lipofection was used as method of transfection in an attempt to increase the number of recovered clones. Surprisingly, a large increase in the proportion of recovered clones to transfected cells was observed using lipofectamine as a method of transfection (280 clones from 7 lipofections transfecting 5.6 x 10^6 cells). Previous reports suggested that there was not much difference in transfection efficiency or toxicity between lipofection and nucleofection in HEK293 or mES cells, however only feeder-

dependent J1 mES cells were tested in that report (Li et al., 2014; Maurisse et al., 2010).

Isolation of correctly targeted knock-in clones

Of the 280 puromycin resistant clones recovered using strategy 4, 145 were positive for PCR screens 2 and 3. Four clones that were positive for two independent PCR screens were expanded and further interrogated. This analysis revealed that two of these clones (L1B7, L1C10) were positive for two additional PCR screens and two (L1C5, L1C6) were positive for three additional PCR screens across the entire insertion cassette (Figure 5.6). The four initial clones (L1B7, L1C5, L1C6, L1C10) were screened further by Southern blot (Figure 5.7). Southern blot screening using a mCherry and 5` probe indicated that a single correct insertion of the *2A-mCherry* cassette in the correct position occurred in three of the clones analysed (L1C5, L1C6, L1C10), and this result was reproducible using freshly isolated DNA.

A triplet band was detected in one of the candidate knock-in clones (L1B7) using the 5` probe which suggested large scale chromosomal rearrangements may have occurred in this clone. Given that the Southern blotting experiments indicated that a duplication event had occurred in this clone during gene targeting, I developed a third screening method to clarify the genotype of the isolated clones. QPCR has often been used to measure copy number variation and genotype for genomic duplication events (Gómez-Rodríguez et al., 2008; Grimholt et al., 2014; Tai et al., 2016). Furthermore, qPCR has successfully been used to detect duplications in the α-globin locus of α-thalassemia patients (Grimholt et al., 2014). I developed a qPCR (Taqman) screening strategy to identify clones without duplication events. I tested the strategy on the candidate *2A-mCherry* clones using a Taqman probe within the 5` probe region that was used in Southern blotting, and two additional probes at the HS-40 enhancer and a region 3` of *HBA2*. According to the Southern blot data, a duplication of the 5` probe region occurred in clone L1B7, while the copy number of this region in clones L1C5, L1C6 and L1C10 should be the same as that of WTH3. As expected, qPCR screening only detected a 1.4-1.7 fold increase in copy number in clone L1B7 with no real difference in copy number between the remaining clones and WTH3. Importantly, qPCR screening detected a 1.7 fold increase in copy

number at the HS-40 probe, located approximately 60 kb upstream of the guide RNA target site, and the 3` probe regions in clone L1B7. This suggested that a chromosomal rearrangement of at least 60 kb had occurred in this clone that resulted in the duplication of the *HBA2* coding sequence and the HS-40 regulatory region. This additional screening by qPCR confirmed that no duplications of the α-globin locus had occurred during gene targeting in the correctly targeted clones L1C5, L1C6 and L1C10.

Phenotypic analysis of 2A-mCherry clones

In vitro differentiation of the candidate *2A-mCherry* knock-in mES clones was performed using the combined method developed in Chapter 3 (pp. 94). Day 7 EBs were analysed by fluorescent microscopy and flow cytometry and all *2A-mCherry* knock-in clones analysed (L1B7, L1C5, L1C6 and L1C10) showed expression of mCherry, although lower levels were detected in clone L1C6 (Figure 5.9; Table 5.3). The activation of mCherry expression in the erythroid lineage validated the PCR and Southern blot screening data which showed a single insertion of the *2A-mCherry* cassette at the *HBA2* locus in these clones (Figure 5.6; Figure 5.7).

Conclusions

In this chapter mES cell clones were isolated that contained a single insertion of the *2A-mCherry* cassette. This was a significant step forward in the generation of a fluorescent reporter system for α-globin in a ΔHS-40 background. As only 8 months remained in this project at the time these cell lines were isolated, it would not have been possible to complete all the remaining gene targeting steps (the removal of the *piggyBac* selection cassette and the deletion of the HS-40 enhancer) and generate mouse colonies from the resulting mES cells for analysis studies. Future work would involve the completion of the remaining gene targeting steps, generation of mouse colonies, isolation of primary red blood cells and ChIP and expression analysis following FACS of the *HBA2-2A-mCherry* expressing and non-expressing populations.

Chapter 6 Epigenetic Editing to Manipulate Bivalent Histone Modifications

6.1 Introduction

Previous chapters in this thesis focused on the generation of a knock-in of a fluorescent reporter in order to efficiently track the transcription of *HBA2* on a single cell level. As previously discussed, *HBA2* transcription is largely controlled by the activity of a distal enhancer. In the absence of this enhancer, erythroid cells express α-globin in a heterogeneous manner, which is correlated with the maintenance of both activating and repressing histone marks in these cell populations (De Gobbi et al., 2017; Vernimmen et al., 2011). The fluorescent knock-in cell line would facilitate the isolation of cells expressing *HBA2* and cells that lack *HBA2* expression, which would then be used to characterise the histone status at the α-globin locus. In this way, the nature of bivalency at the α-globin locus would be assessed, however, these experiments would not be informative as to the function of this bivalency. One way to understand how bivalency influences gene expression would be to analyse gene expression changes during differentiation after depleting or accumulating histone marks at a specific gene. In this manner, a causal relationship between bivalency resolution, gene expression and development can be established.

In this chapter, I describe the development of a strategy aimed towards modifying the presence of chemical groups on the tails of histones, which are associated with gene expression behaviours. Initially, I developed a high-throughput and efficient cloning system that facilitates the construction of gene fusions containing a combination of promoters, DNA binding domains and enzymatic domains. Vectors containing these gene fusions were transfected into cells in order to assess their effects on a model locus (Figure 6.1). Before presenting the results of this study, I briefly discuss the experimental design below.

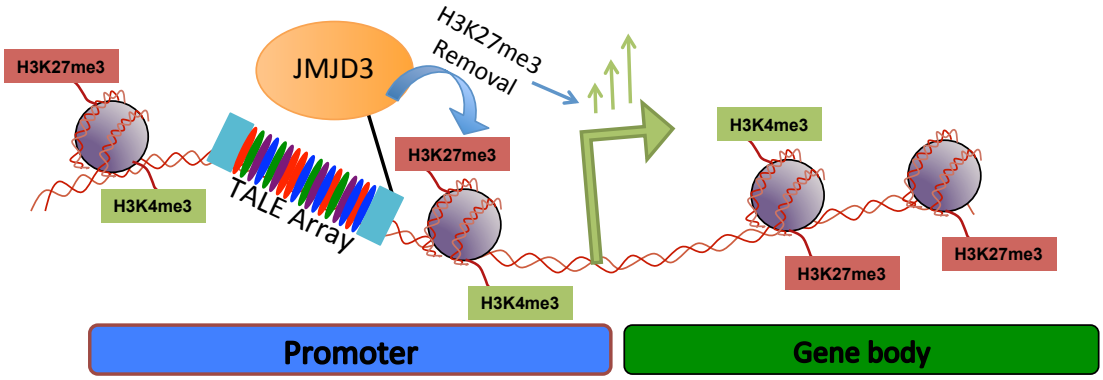


Figure 6.1 Using TALE fusion protein activity to modify histone status.
A TALE array can be designed to interact with any chosen DNA sequence. If the TALE array is fused to an epigenetic editing enzymatic domain, modifications of the histone state in the vicinity of the binding event may occur. In the example shown here, JMJD3, a H3K27me3 demethylase, is targeted to a promoter marked with H3K27me3 and H3K4me3. The removal of H3K27me3 at the bivalent gene promoter by JMJD3 may be sufficient to increase the expression of the gene. This figure is adapted from Vernimmen (2014).

Golden Gate cloning to generate TALE fusion proteins

Both the catalytically inactive Cas9 (dCas9) coding region and TALEs can be fused to the enzymatic domains of other proteins (Hilton et al., 2015). Here, I used TALE fusions to direct site-specific epigenetic changes, as at the time these experiments were designed, these artificial DNA-binding domains had already been validated for this purpose (Konermann et al., 2013; Maeder et al., 2013a; Mendenhall et al., 2013). A modified version of the Golden Gate assembly protocols used in Cermak *et al.* (2011) was used to construct the TALE fusions in this chapter because of the flexibility and efficiency this technique offers. Golden Gate assembly is a cloning method based on the use of type IIS restriction enzymes, which cut DNA close to, but outside, their continuous, asymmetric recognition sites (Figure 6.2A). These enzymes were first used in cloning schemes in 1996 (Lee et al., 1996; Padgett and Sorge, 1996) where they were employed to generate multimers of individual cloned fragments of DNA. When cloning schemes are appropriately designed, type IIS restriction enzymes can be used to generate combinations of unique and

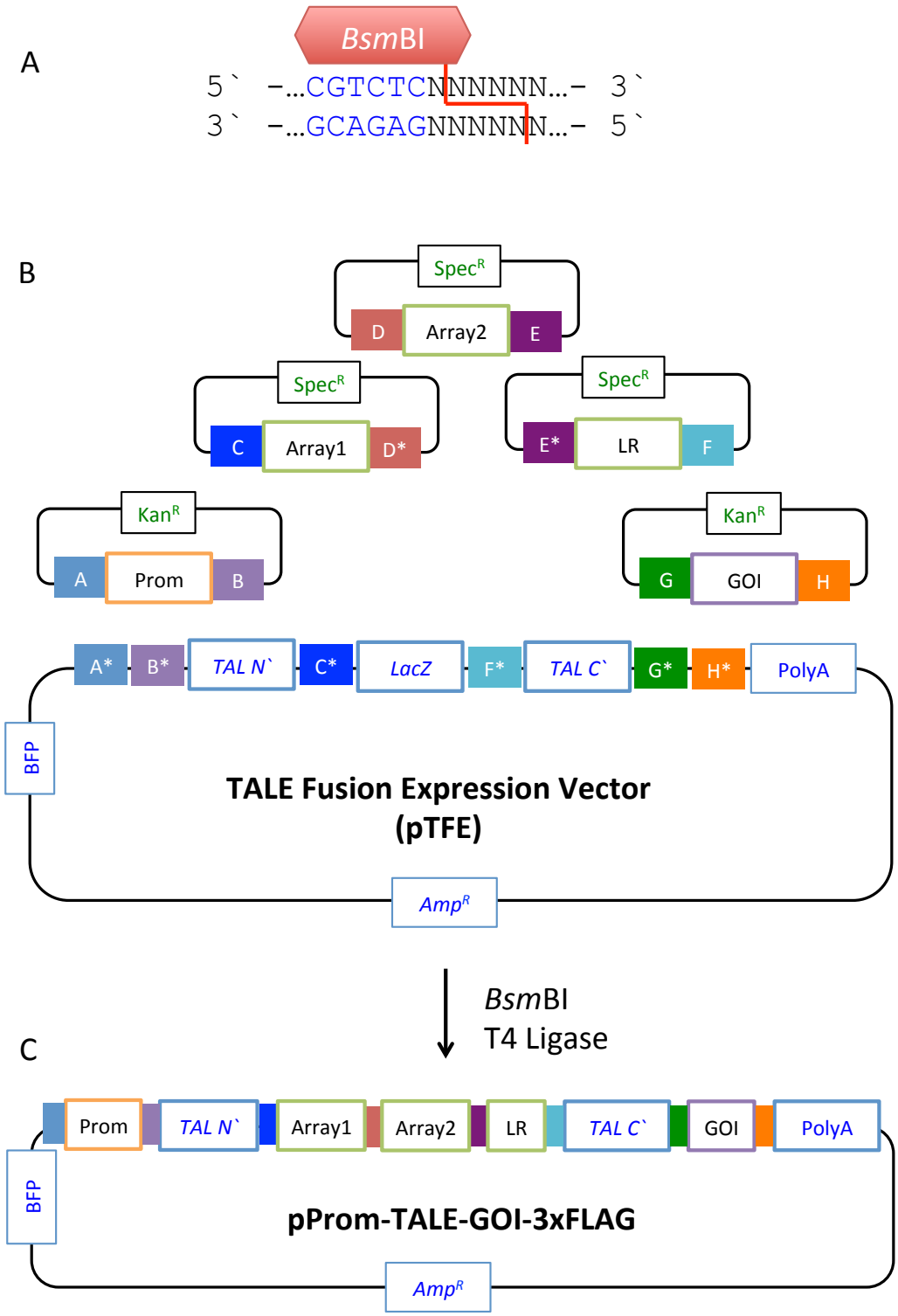


Figure 6.2 Golden Gate assembly of TALE constructs.
A. The recognition site of the type IIS restriction enzyme *Bsm*BI (and its isochizomer, *Esp*3I) is shown in blue. This enzyme cleaves its recognition site at the position shown as a red line. *Figure legend is continued on the next page.*

complementary cohesive ends on different DNA modules, which can then be ligated together in a particular order (Figure 6.2B). This method of using type IIS restriction sites to assemble at least nine DNA modules together was termed 'Golden Gate' cloning (Engler et al., 2008, 2009). Later methods based on Golden Gate cloning, including MoClo (Modular Cloning) and GoldenBraid, developed libraries of plasmids containing DNA modules which could be shuffled to assemble a vast number of final constructs with minimal effort (Sarrion-Perdigones et al., 2011; Weber et al., 2011). For example, a library of Golden Gate compatible plasmids containing different promoters, gene coding regions and terminators can be shuffled to produce a large number of combinations in the final expression vector.

Figure 6.2 Legend continued:
(B-C). Golden Gate cloning strategy for fusing five DNA modules from plasmids into the destination vector, TALE Fusion Expression Vector (pTFE). Before initiating the Golden Gate assembly protocol to generate the Promoter-TALE-enzymatic domain cassette, several entry plasmids were prepared. Two separate plasmids that contain the 5` and 3` (Array 1 and Array 2) section of the complete TALE array were cloned while the plasmid containing the last half repeat (LR) is available as part of the TALE assembly kit (Cermak et al., 2011). In addition, two other plasmids harbouring the promoter and enzymatic domains, respectively, were cloned. Each of these regions of interest are flanked by Type IIS restriction sites, which were designed to re-ligate in a specific order upon addition of the restriction enzyme and ligase. The TALE array (Array 1, Array 2 and LR) was designed to insert between the TAL N- and C- terminal domain coding regions. The promoter and enzymatic domain DNA sequences were designed to insert up- and downstream, respectively, of the TAL N- and C- terminal sequences. Digestion with *Bsm*BI releases the insert modules, each flanked with a unique 4 nt overhang (colored boxes A-H). These overhangs anneal to their complementary cohesive ends (colored boxes A*-H*) and are ligated to one another in a specific order in the presence of T4 Ligase. The selection of colonies was based on ampicillin resistance of pTFE while blue-white selection was also employed, as the *LacZ* cassette is lost once the TALE array has been inserted into pTFE. TAL N` and C`, TAL N- and C-terminal domains; BFP, CMV-mTagBFP-polyA cassette; LacZ, a LacZ cassette; Amp^R, an ampicillin resistant cassette; Prom, promoter of choice; GOI, gene of interest; 3xFLAG, three times FLAG tag; Array 1, RVD 1-10; Array 2, RVD 11-(n-1); LR, the last half repeat RVD n; Kan^R, kanamycin resistance cassette; SpecR, spectinomycin resistance cassette.

The protocols described in Cermak *et al.* (2011) use Golden Gate cloning to sequentially assemble TALE DNA binding arrays into a destination vector. A TALE array can be reliably designed to target a specific genomic locus using online tools (Doyle et al., 2012). Once a TALE array containing ~12 – 30 RVDs has been designed it is then assembled in two steps (Cermak et al., 2011). First, the RVD modules 1-10 and 11-(n-1) (Array 1 and 2) are assembled into two spectinomycin resistant plasmids, respectively, using Golden Gate assembly, where n is the number of RVDs in the final TALE array. Arrays 1, 2 and the last half repeat are then inserted into a destination vector in between the N- and C-terminal coding TALE domains using Golden Gate Assembly. I modified the pTal2 destination vector (Cermak et al., 2011), to generate a TALE Fusion Expression vector, pTFE, by introducing additional Golden Gate restriction sites, a CMV-mTagBFP-polyA cassette (BFP) and a poly-A signal (Figure 6.2B). The cloning scheme was designed such that each of the sequences encoding enzymatic domains are separated from the coding regions of the TALEs by a sequence encoding a short Glycine-Glycine-Glycine-Serine linker in the final fusion constructs (previously used in TALE fusions in (Maeder et al., 2013a)). The inclusion of this linker region was intended to prevent potentially inhibitory interactions between the enzymatic domains and the TALE proteins. The subcloning vectors were designed such that each of the genes would be fused to sequences encoding a 3xFLAG tag at the 3` end, which is a commonly used epitope tag (3xFLAG has previously used in epigenetic editing fusion constructs in (Hilton et al., 2015; Kearns et al., 2015; Maeder et al., 2013a; Mendenhall et al., 2013)). The addition of these linker and epitope sequences was designed to preserve the reading frames of the final fusion genes.

Using Nrp1 as a model target to test epigenetic editing tools

NRP1 is a transmembrane receptor that binds vascular endothelial growth factor (VEGF) (Fantin et al., 2013; Kawasaki et al., 1999). NRP1 is expressed in developing neurons and is important during axon guidance, vascular sprouting and brain angiogenesis (Aspalter et al., 2015; Fantin et al., 2013; Kawasaki et al., 1999). The *Nrp1* promoter contains a CpG island, is bivalent (Figure 6.3A) and does not

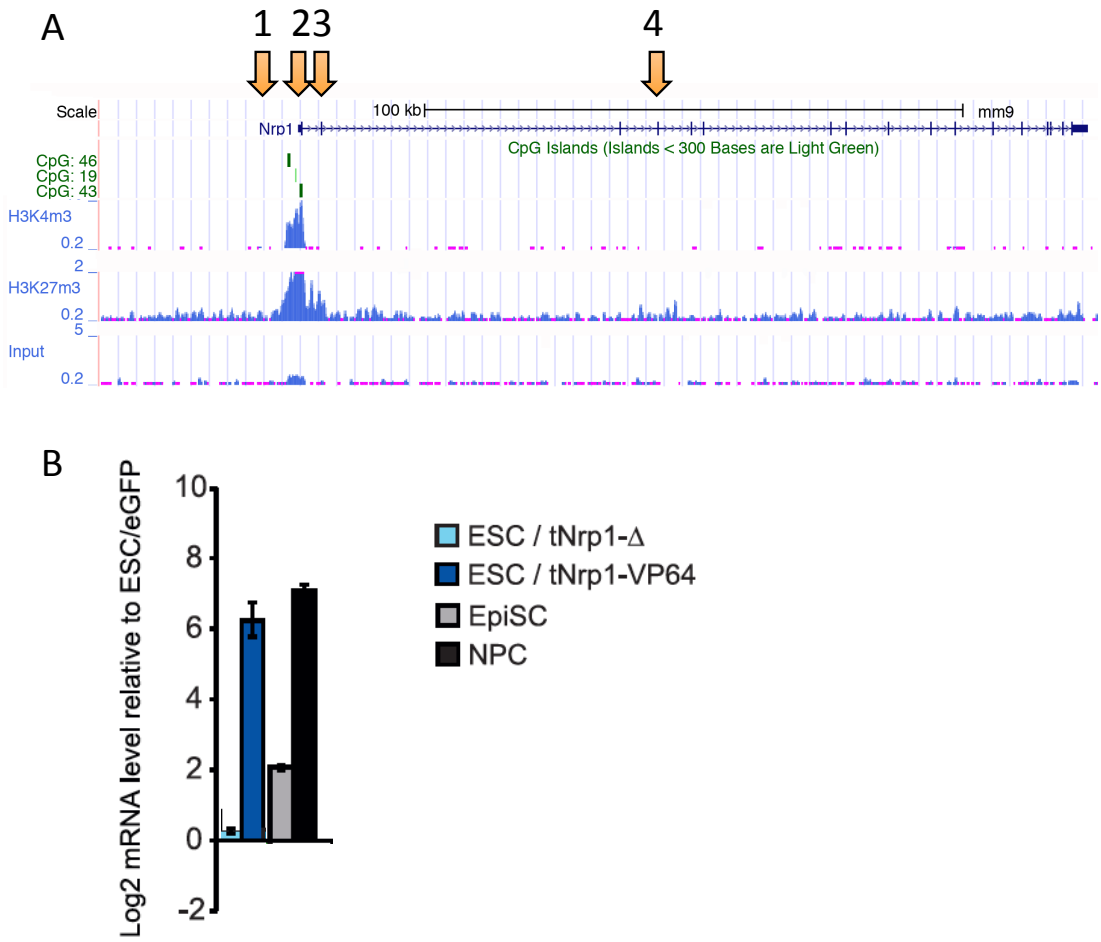


Figure 6.3 *Nrp1* characteristics in mES cells.
A. Enrichment tracks for H3K4me3 and H3K27me3 in "Bruce4" mES cells at *Nrp1* are shown from ChIP sequencing data of histone modifications from ENCODE/LICR data set release 3 (August 2012) and mapped on the mouse July 2007 (NCBI37/mm9) genome assembly (http://genome.ucsc.edu/; Kent et al. 2002). The locations of four primer pairs used in ChIP-qPCR in this chapter are shown as orange arrows.
B. RT-qPCR data for *Nrp1* in mES cells transfected with the TALE vector tNrp1-VP64 and the negative control TALE tNrp1-Δ. The expression shown is relative to *GAPDH* expression and to eGFP transfection. Expression changes are also shown for mES cells that have differentiated into epiblast stem cells (EpiSCs; by addition of growth factors activin A and Fgf2) or NPCs (by retinoic acid mediated differentiation). The mean $\log_2$ expression data from three biological replicates is shown. This figure is adapted from Therizols *et al.*, (2014).

bear DNA methylation in mES cells (Sabo et al., 2004, 2006). *Nrp1* expression can be detected at low levels in mES cells and increases as cells differentiate towards the neural lineage (Mikkelsen et al., 2007; Therizols et al., 2014). Therizols *et al*. (2014) from the Bickmore lab at the University of Edinburgh designed a TALE that targeted the *Nrp1* promoter and was fused to the strong acidic transactivation domain, VP64 (tNrp1-VP64; Therizols et al. 2014). This TALE fusion promoted efficient and robust activation of *Nrp1* in mES cells transiently transfected with tNrp1-VP64 to levels comparable with the endogenous expression of *Nrp1* in neural precursor cells (NPCs; Figure 6.3B). Because the *Nrp1* locus bears similarities to the *HBA2* locus (bivalent, CpG island-containing and methylation-free DNA) and because of the availability of validated tools (tNrp1 and tNrp1-VP64), *Nrp1* was selected as a model target site for the epigenetic editing tools generated in this chapter. The validated TALE sequence targeting the *Nrp1* promoter was used to generate the TALE fusion constructs while plasmids from the Bickmore lab containing the tNrp1-VP64 coding regions and the tNrp1 coding region alone (tNrp1-Δ, TALE-Only-B) were used as positive and negative controls, respectively (Therizols et al. 2014).

Using enzymatic domains to manipulate histone modifications

The relationship between histone status and gene expression is unclear. H3K4me3 and H3K27me3 are commonly described as an "activator" and a "repressor" of gene transcription, respectively. As described in Chapter 1 (pp. 2), a number of studies have shown that a dramatic global reduction in H3K4me3 does not result in major changes in global gene expression levels (Clouaire et al., 2012; Hormanseder et al., 2017; Lenstra et al., 2011; Muramoto et al., 2010). Also, H3K27me3 deposition has been shown to occur only after transcription has been silenced (Hosogane et al., 2013; Yuan et al., 2012). The instructive role of these two histone marks in gene transcription is in doubt, as indeed is the importance of colocalisation of these two marks for appropriate gene expression during differentiation (Hu et al., 2013). These marks may, however, play an important role in terms of epigenetic memory of transcriptional states (Coleman and Struhl, 2017; Hormanseder et al., 2017; Howe et al., 2017; Muramoto et al., 2010). To improve our

understanding of these histone marks, I aimed to promote methylation of H3K4 and H3K27, respectively, and demethylation of H3K27.

A number of proteins are involved in catalysing H3K4me3 deposition (see Chapter 1, pp. 2). The RBBP5 protein is a core member of the SET1-like MLL1 complex and, along with the other two subcomponents, WDR5 and ASH2L, it is essential for the assembly and H3K4me3 methyltransferase activity of the SET1-like MLL complex (Avdic et al., 2011; Cao et al., 2010; Dou et al., 2006). The catalytic components of the SET1 and SET1-like complexes are often extremely large, ranging from ~5-17 kb in gene length and ~200-600 kDa in protein size. These large DNA regions may be difficult to clone and the large protein domains may interfere with TALE binding activity. Therefore, the small *Rbbp5* coding region (~1.6 kb) was cloned and tested. Localising the RBBP5 protein to a particular locus may stimulate the recruitment of the entire MLL complex, thus stimulating H3K4me3 deposition and, perhaps, transcription. This hypothesis is supported by the observation that targeted high levels of H3K4me3 deposition by the meiotic methlytransferase PRDM9 is sufficient to induce small changes in transcription at silent loci in a context dependent manner (Cano-Rodriguez et al., 2016).

EZH2 is a core member of the H3K27me3 histone methyltransferase complex, PRC2 (O' Carroll et al., 2001). Although EZH2 is the catalytically active component of the PRC2 complex, the presence of both SUZ12 and EED are required to stimulate this activity (Margueron et al., 2009; Pasini et al., 2004). Previous reports suggest that inducing H3K27me3 at specific promoters significantly reduces gene expression (Konermann et al., 2013). As discussed in detail in the Introduction (pp. 42), a recent report using three sgRNAs targeting dCas9-EZH2 to a gene promoter was able to induce a 9-fold enrichment of H3K27me3 at the gene promoter, however, the effect of this localisation on transcription was unclear (O'Geen et al., 2017). It is possible that the localisation of the EZH2 protein to a particular locus can stimulate the recruitment of the entire PRC2 complex and thus stimulate H3K27me3 deposition and perhaps suppress transcription.

JMJD3 is a H3K27me3 demethylase important in transcriptional regulation and development (Agger et al., 2007; Burgold et al., 2008; Hong et al., 2007). A C-

terminal segment of JMJD3, including the conserved JmjC domain, is sufficient for its demethylase activity *in vitro* (Kruidenier et al., 2012). The demethylase activity of JMJD3 at some bivalent loci is important for PolII release, which allows for complete transcription elongation (Chen et al., 2012). Targeting the JMJD3 protein to a particular locus may be sufficient to remove H3K27me3 marks and promote transcription.

A tNrp1-VP64 fusion protein was used as a positive control to confirm the induction of *Nrp1* expression *in vitro* by TALE targeting. As discussed in Chapter 1 (pp. 40), the level of induction of gene expression by different strategies using VP64 constructs can vary substantially. Notably, a 100,000 fold induction in target gene expression only causes a 5-20 fold enrichment of H3K4me3 (Black et al., 2016). However, an ~20x induction of target gene expression is not sufficient to induce H3K4me3 enrichment at the gene promoter (Cano-Rodriguez et al., 2016). In this chapter, I test the effects of a tNrp1-VP64 fusion on *Nrp1* gene expression and H3K4me3 enrichment.

Here, I will describe the generation of a number of Golden Gate subcloning vectors containing a variety of DNA modules. *Nrp1* was used as a model target gene with which to test this system, however, the intention was to translate the validated methodology to target the human α-globin locus in humanised mES cells. Therefore, this system was optimised in the humanised mES cell line, WTH3, which is used in all the transfection experiments described. Vectors were produced containing DNA modules whose protein products are designed to modify the presence of methyl groups on different histone tails associated with the *Nrp1*. I will describe the effects of transfecting WTH3 cells with these vectors, which was assessed by qRT-PCR and ChIP-qPCR.

6.2 Results

6.2.1 Generation of a TALE fusion expression vector cloning system

In this section I describe the design and generation of a TALE fusion expression vector cloning system (Figure 6.4). The system was based on the Golden Gate assembly protocols and plasmids used to generate TALE array constructs in Cermak *et al.* (2011). To generate an expression vector into which a promoter, a TALE array and a gene of interest (GOI) could be cloned, I first modified the pTal2 vector (Cermak et al., 2011). I cloned additional *Bsm*BI Golden Gate restriction sites into pTal2 to facilitate the addition of a promoter and GOI into this vector by the Golden Gate methodology. I also inserted a CMV-mTagBFP-polyA cassette, to allow for the identification of transfected cells by fluorescent microscopy, and a poly-A signal downstream of the entry site for the GOI. The final ampicillin resistant expression vector, pTALE-Fusion-Expression vector (pTFE) was used in Golden Gate restriction reactions (Figure 6.2B, Figure 6.4A).

I then generated kanamycin resistant subcloning vectors that were used in Golden Gate reactions that contained the GOI tagged with 3xFLAG, and the promoter of choice. The epigenetic editing genes used in these TALE fusions were first cloned into pUC19 vectors (Figure 6.4B). Two kanamycin resistant cloning vectors, pKan^{R}-MCS (pStb205 and pGreen) were then equipped as GOI entry vectors (pKan^{R}-MCS-3xFLAG-GG) by the cloning of additional restriction sites and a 3xFLAG sequence (Figure 6.4C). The sequences encoding the epigenetic editing enzymes were then subcloned into one of the pKan^{R}-MCS-3xFLAG-GG vectors to generate kanamycin resistant vectors appropriate for Golden Gate assembly (Figure 6.4D). A phosphoglycerate kinase (PGK) promoter was also inserted into the kanamycin resistant vector, pGreen, with the addition of appropriate Golden Gate restriction sites (Figure 6.4E). The TALE array plasmids were then assembled (Figure 6.4F) as described previously (Cermak et al., 2011).

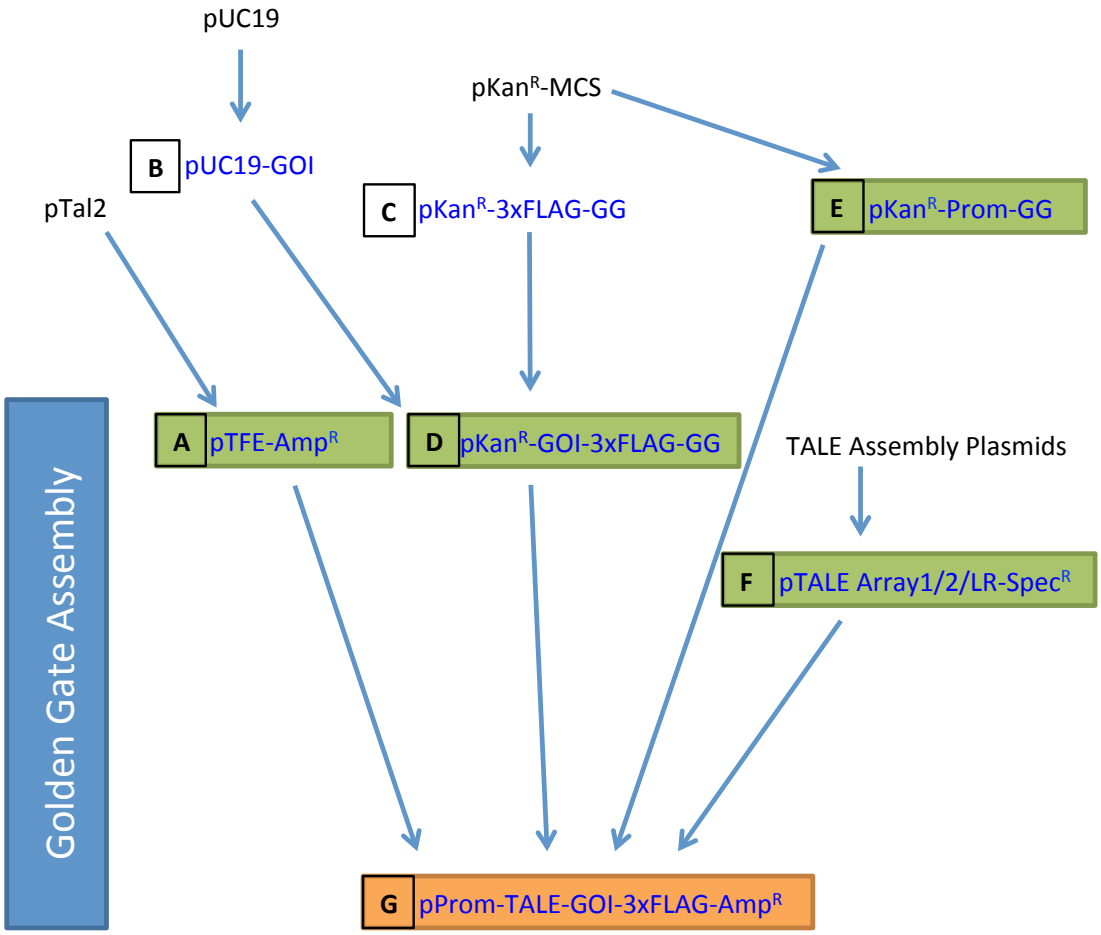


Figure 6.4 Schematic showing cloning steps for generating expression plasmids for fusing epigenetic editing genes of interest (GOI) to site-specific TALE arrays. Plasmid construct groups shown in blue text were cloned in this study. Green boxes indicate the constructs used in the final Golden Gate assembly reactions to generate the TALE fusion expression vectors (orange box) used in transfections later in this chapter. The details of the cloning scheme (A-G) are described in the text (section 6.2.1). The pTal2 plasmid and the TALE Assembly plasmid kit (Cermak et al., 2011) were received as a gift from Dr. Peter Hohenstein (University of Edinburgh). The pKan^R-MCS plasmids pStb205 and pGreen were received as a gift from Dr. Diarmuid Ó Maoiléidigh (MPIPZ, Köln).

Finally, the completed fusion expression vectors were assembled (Figure 6.4G). This was achieved by the ligation of *Bsm*BI digested Golden Gate subcloning vectors (green boxes in Figure 6.4). These vectors contained a TALE array designed to target the *Nrp1* promoter fused to a GOI-3xFLAG under the control of a promoter of choice.

6.2.1.1 Cloning epigenetic editing genes from cDNA into pUC19 vectors

Three genes were cloned from WTH mES cell cDNA libraries using the scheme outlined in Figure 6.5. Primers were designed to amplify the full-length mouse coding region of *Rbbp5* (1,617 bp) and *Ezh2* (2,241 bp). Primers were also designed to amplify the C-terminal coding region of mouse *Jmjd3* (1,467 bp), including the sequences that encode the amino acid regions (1130-1641) required for catalytic activity (Kruidenier et al., 2012). The fragments of interest were isolated after separation on an agarose gel and phosphorylated using T4 PNK. The pUC19 vector was linearised by digestion with the blunt end restriction enzyme *Sma*I and dephosphorylated with CIP. The purified insert fragments were then ligated into a purified vector fragment. Ligated products were transformed into chemically competent *E. coli* and successful cloning of resulting plasmids (Figure 6.6A) was confirmed by restriction enzyme pattern analysis with *Eco*RI and *Hind*III (Figure 6.6B). The entire cloned fragment was sequenced using Sanger technology and although a number of SNPs were identified in the cloned *Rbbp5* and *Jmjd3* fragments no non-synonymous SNPs were identified in any of the cloned genes (Table 6.1).

Table 6.1 SNPs identified within cloned genes. SNPs were identified in the *Rbbp5* and *Jmjd3* genes cloned from WTH3 mES cells by comparison of Sanger sequencing data from cloned vectors to the published reference sequence for mouse on NCBI. All SNPs identified were previously annotated in dbSNP (NCBI) and are synonymous mutations that do not change the identity of the amino acid residue. No SNPs were detected in the cloned *Ezh2* sequence.

Gene	SNP	refSNP	NCBI ref seq	WTH cDNA seq	Residue change
RBBP5	1	rs31916079	A	G	Gly - Gly
	2	rs6350018	G	C	Ser - Ser
	3	rs3700112	A	G	Lys - Lys
	4	rs32724155	T	C	His - His
JMJD3	1	rs13467325	A	G	Val - Val

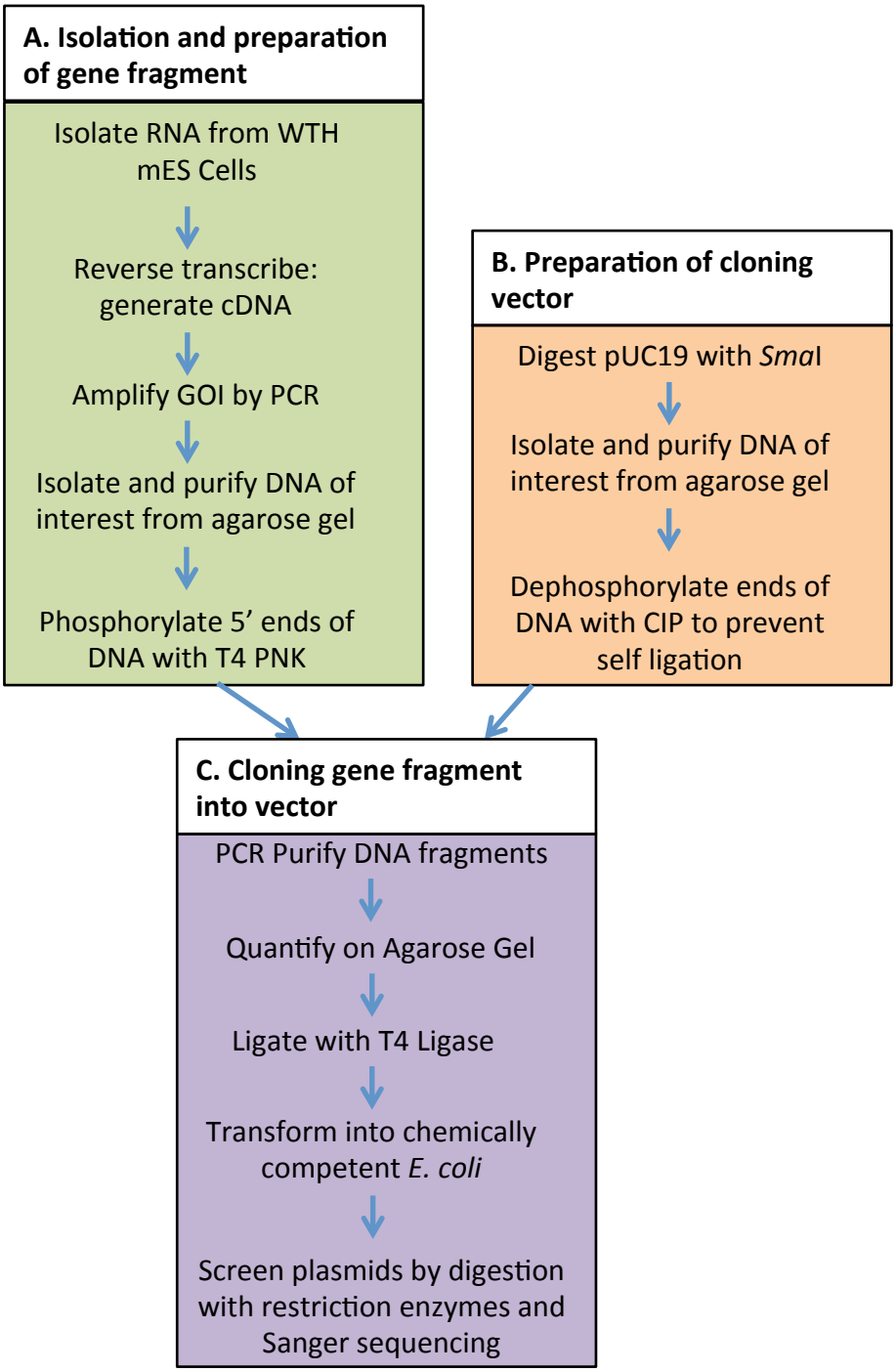


Figure 6.5 Schematic showing steps for cloning epigenetic editing genes.
A. Genes of interest (GOI; *Rbbp5*, *Ezh2* and *Jmjd3*) were amplified by PCR from cDNA libraries prepared using WTH mES cell RNA as a template. Gene fragments were isolated and purified on agarose gels and then phosphorylated with T4 PNK.
B. The pUC19 cloning vector was linearised by digestion with *Sma*I. The vector was then isolated and purified on an agarose gel and dephosphorylated with CIP.
C. The GOI and vector fragments were purified, quantified, ligated and then transformed into *E. coli*. Plasmids were isolated from *E. coli* colonies and screened by restriction digestion and Sanger sequencing.

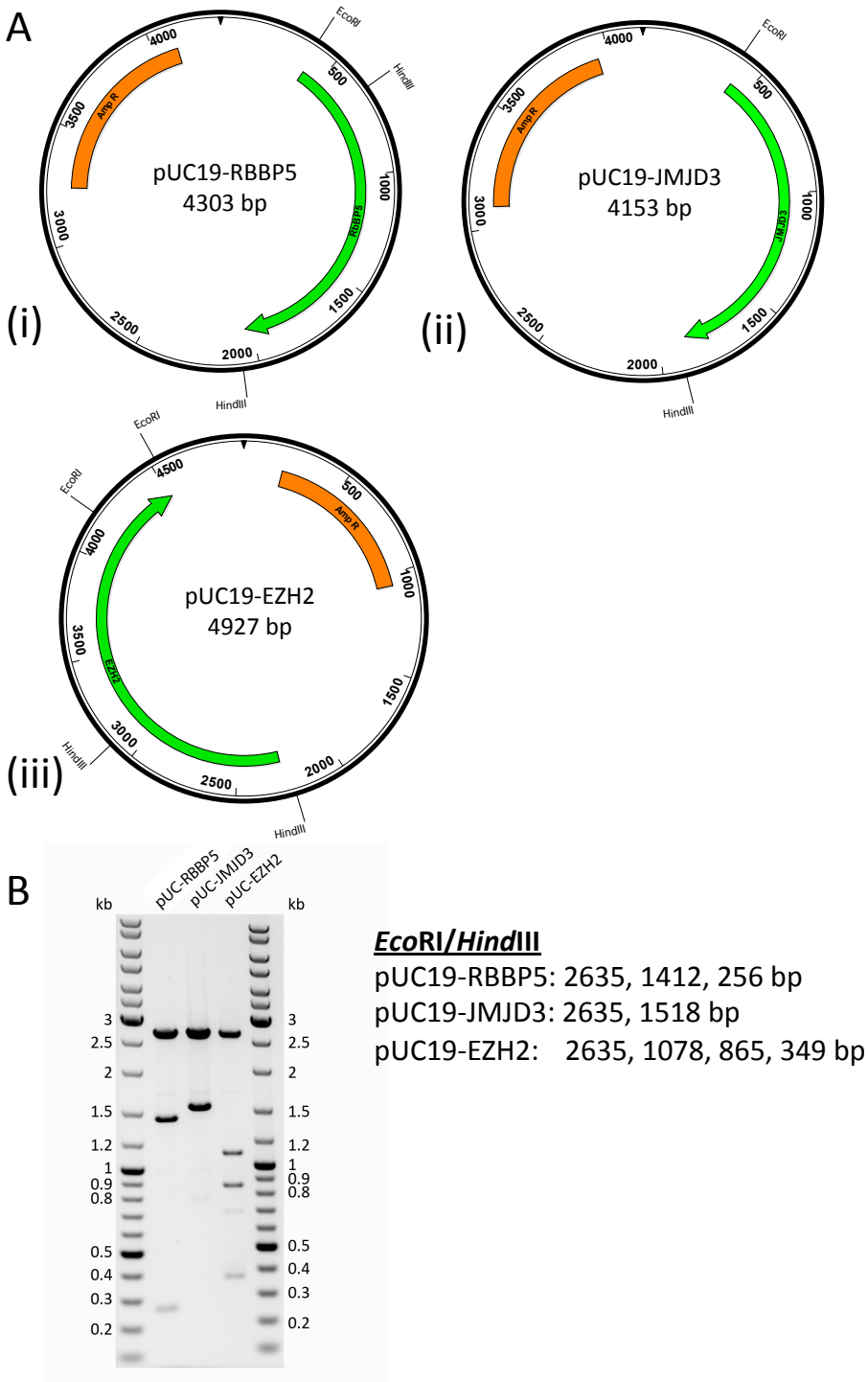


Figure 6.6 Cloning epigenetic editing genes from cDNA into pUC19.
A. Plasmid maps of **(i)** pUC19-RBBP5, **(ii)** pUC19-JMJD3 and **(iii)** pUC19-EZH2.
B. Agarose gel images showing restriction digestion with *Eco*RI/*Hind*III of three plasmids confirming cDNA from epigenetic editing genes has been successfully cloned into the pUC19 vector. The expected restriction digestion patterns are shown.

6.2.1.2 Generation of FLAG-tagged subcloning vectors

TALE fusion constructs were generated using a modified protocol from Cermak *et al.* (2011), which employs Golden Gate cloning using plasmids with counter selective antibiotic resistant cassettes to increase screening efficiency. Here, I modified two available plasmids to generate vectors that included 3xFLAG tags followed by a TAA stop codon and additional restriction enzyme sites (Figure 6.7A). These plasmids harbour cassettes that confer kanamycin resistance to bacteria, which can be counter-selected against the ampicillin-resistant phenotypes of pTal2-containing bacteria. To generate these plasmids, complementary single-stranded oligonucleotides containing the 3xFLAG cassette were synthesised (Sigma), annealed and digested with the restriction enzymes *Xba*I and *Not*I. The digested fragment was ligated with the pStb205 and pGreen vectors, which were also digested with *Xba*I and *Not*I and dephosphorylated with the enzyme rSAP (Figure 6.7A). These ligation reactions were transformed into chemically competent *E. coli* and plated on LB/agar supplemented with kanamycin. The successful cloning of the resulting plasmids, pStb205_3xFLAG and pGreen_3xFLAG (Figure 6.7B ii and iv) was confirmed by digestion with restriction enzyme *Bsm*BI (Figure 6.7C) and Sanger sequencing (data not shown).

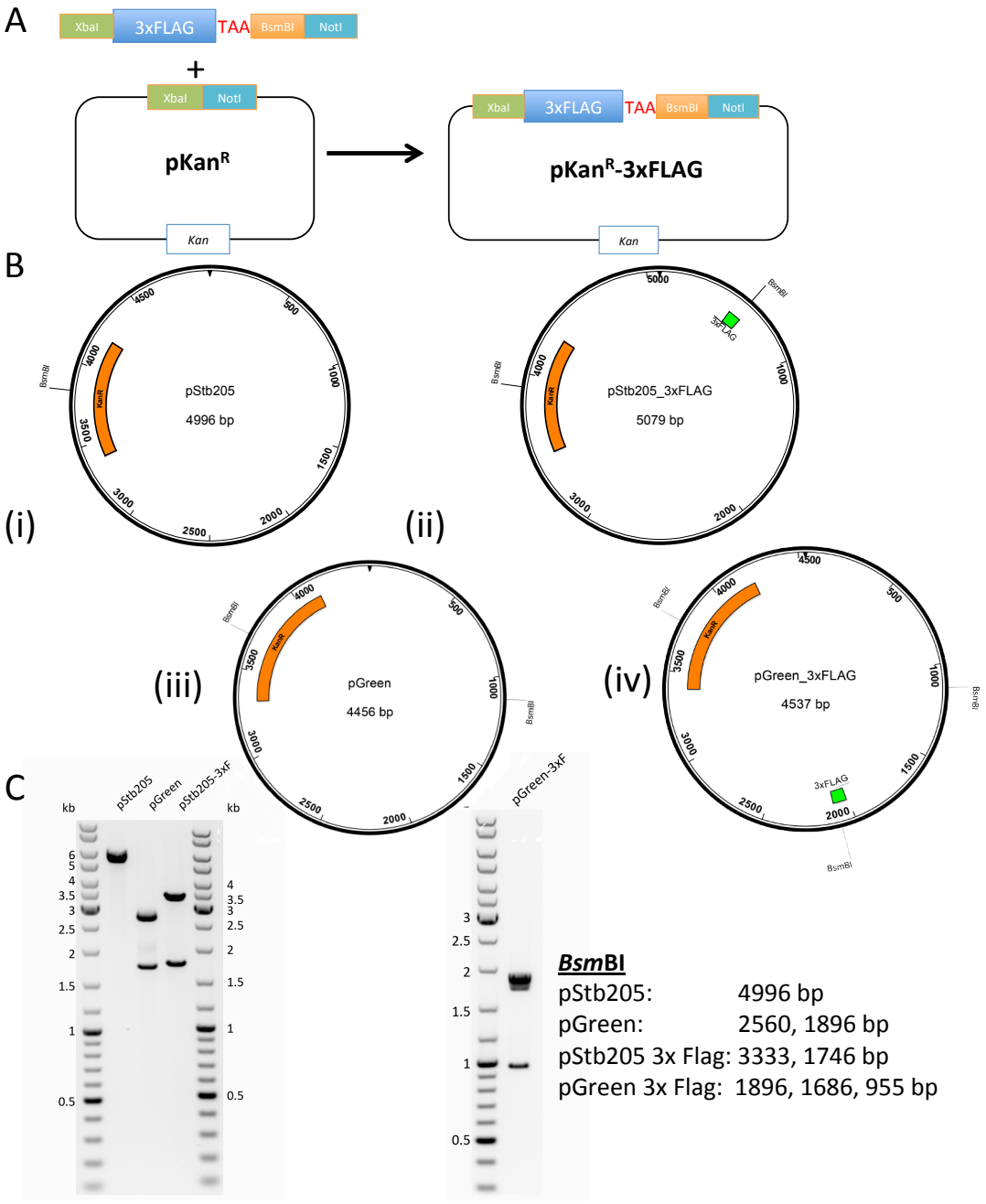


Figure 6.7 Cloning 3x FLAG into Kan^R vectors to generate Kan^R Golden Gate subcloning vectors for epitope tagging genes of interest.
A. Cloning strategy to generate Kan^R subcloning vectors including a 3xFLAG tag and additional restriction sites. A synthesised oligonucleotide containing restriction sites and a 3xFLAG tag was cloned into two Kan^R vectors; pStb205 and pGreen.
B. Plasmid maps of **(i)** pStb205, **(ii)** pStb205_3xFLAG, **(iii)** pGreen and **(iv)** pGreen_3xFLAG.
C. Gel image showing restriction digestion with *Bsm*BI of plasmids in B confirming the expected digestion patterns of the 3xFLAG-containing plasmids was observed. The expected restriction digestion patterns are shown.

6.2.1.3 Generation of a subcloning vector containing the PGK promoter

The PGK constitutive promoter was also cloned into the kanamycin resistant vector, pGreen, using a similar strategy to that shown in (Figure 6.7A). Briefly, primers were designed to amplify the PGK promoter from pAAT:PGK:PB:Puro-delta-TK (Yusa et al., 2011) and include additional restriction sites. The PCR product was purified and phosphorylated with the enzyme T4 PNK. The pGreen vector was digested with *Not*I and *Xba*I and the resulting linearised vector was dephosphorylated with the enzyme rSAP. The concentrations of the purified vector and insert fragments were quantified on an agarose gel and ligated overnight. Ligated products were transformed into chemically competent *E. coli* and plated on LB/agar supplemented with kanamycin. Resulting colonies were screened by colony PCR (data not shown) and correct cloning of the pGreen-PGK vector (Figure 6.8A) was confirmed by restriction digestion with *Bsm*BI (Figure 6.8B) and Sanger sequencing (data not shown). Although a *Bsm*BI site is located within the PGK promoter this site generates overhangs that are not compatible with any of the overhangs designed for Golden Gate cloning in this cloning strategy (see Appendix, pp. 249). As such, the PGK promoter will be digested by *Bsm*BI but this site should reconstitute itself following ligation.

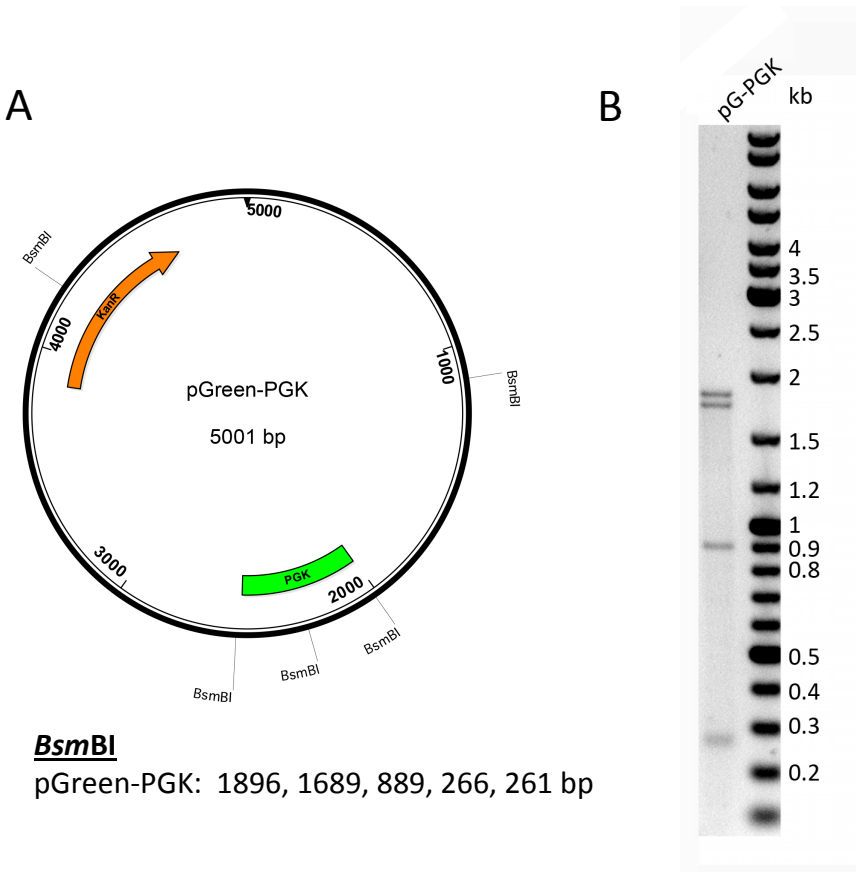


Figure 6.8 Cloning the PGK promoter into Kan^R vectors to generate a Kan^R Golden Gate subcloning vector.
A. Plasmid map of pGreen-PGK.
B. Gel image showing restriction digestion of pGreen-PGK with *Bsm*BI confirming the PGK promoter was successfully cloned. The expected restriction digestion pattern is shown.

6.2.1.4 Cloning epigenetic editing genes and promoter into Golden Gate subcloning vectors

In order to generate GOI subcloning vectors that could be used in Golden Gate restriction reactions, the GOIs were cloned into the Kan^R-3xFLAG vectors in a manner that generated additional Golden Gate restriction sites. The epigenetic editing genes *Rbbp5*, *Ezh2* and *Jmjd3* were cloned from pUC19 vectors (Figure 6.6A) into kanamycin resistant subcloning vectors, pStb205_3xFlag or pGreen_3xFlag (Figure 6.7Bii and iv, respectively) using the strategy outlined in Figure 6.9. Briefly, primers were designed that included additional restriction digest sites to amplify the genes of interest. Following amplification, the PCR products were digested with the restriction enzymes *Eco*RI and *Xba*I. The digested products were isolated from an agarose gel and phosphorylated with the enzyme T4 PNK. The pStb205_3xFlag vector was also digested with *Eco*RI and *Xba*I and was dephosphorylated using the enzyme rSAP. The insert and vector fragments were purified and ligated overnight. The ligation reactions were transformed into chemically competent *E. coli* and plated on LB/agar supplemented with kanamycin. The resulting colonies were screened by PCR (data not shown) and the successful cloning was confirmed by restriction digestion with *Bsm*BI (Figure 6.10B) and Sanger sequencing (data not shown).

This strategy proved successful in cloning the plasmids pStb205_RBBP5_3xFlag and pStb205_EZH2_3xFlag (Figure 6.10Ai and iii). However, no clones were isolated containing the complete, correctly orientated *Jmjd3* sequence from pUC19-JMJD3. The same cloning strategy (Figure 6.9) was used to clone *Jmjd3* into the pGreen_3xFlag vector, which did not contain Gateway recombination sites. This generated the plasmid pGreen_JMJD3_3xFlag (Figure 6.10Aii), which was confirmed to be correct by restriction digestion with *Bsm*BI (Figure 6.10B) and Sanger sequencing (data not shown).

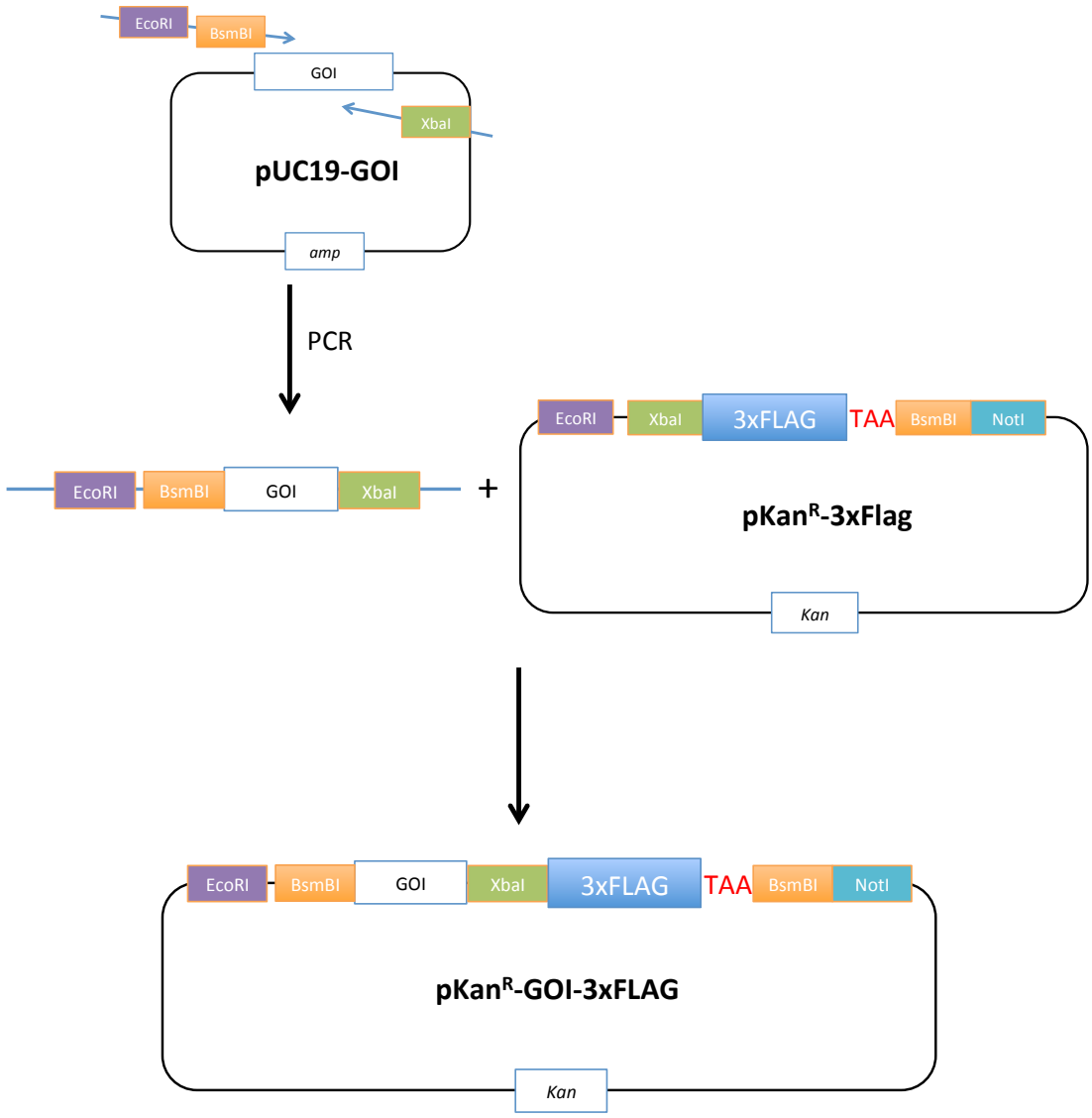
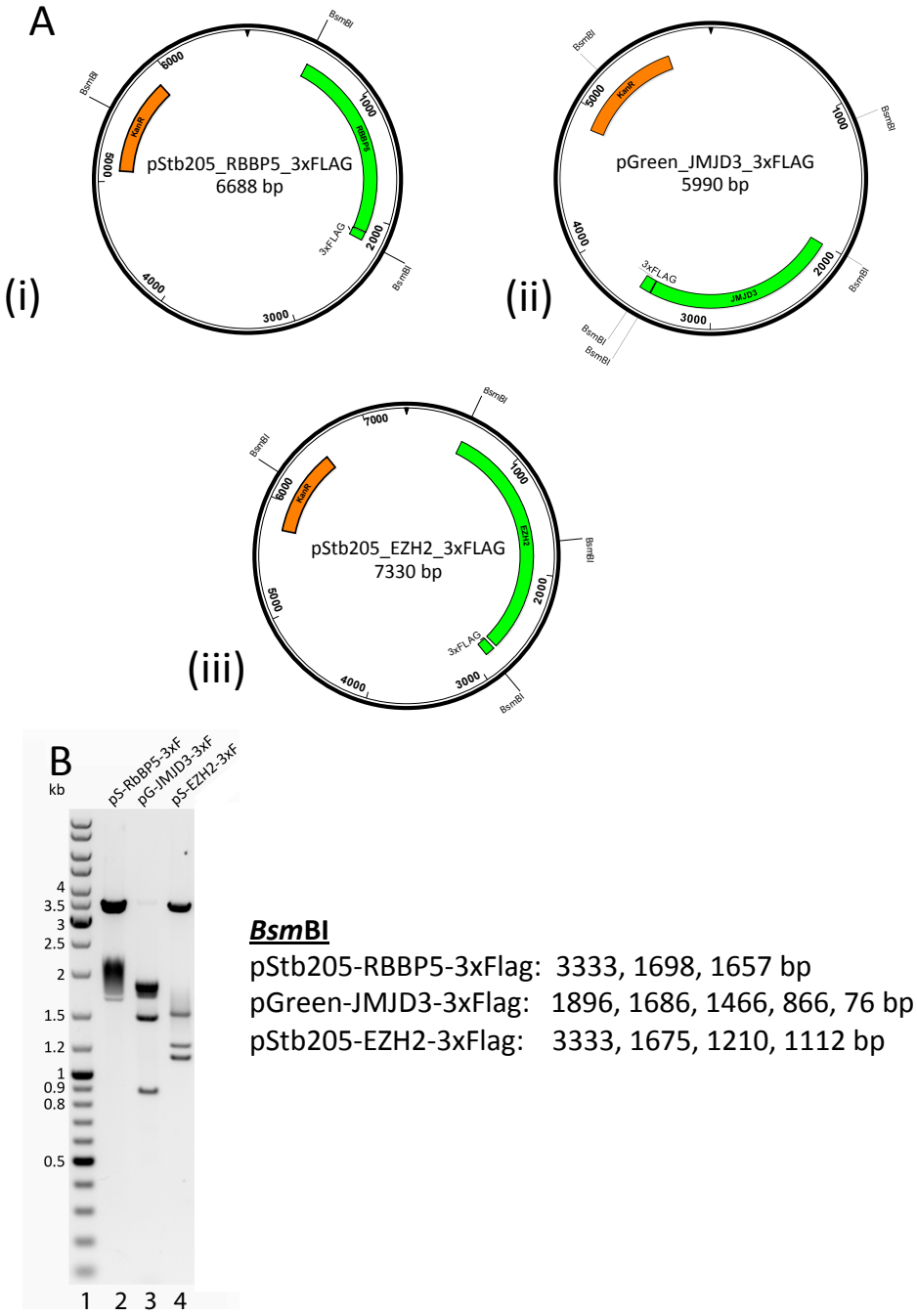


Figure 6.9 Cloning strategy to generate Kan^R subcloning vectors including a gene of interest (GOI) fused to 3xFLAG tag.
The genes of interest were amplified by PCR from plasmids using primers that contained additional restriction sites. The resulting PCR products were cloned into a Kan^R subcloning vectors that contained a 3xFLAG tag.



***Bsm*BI**

pStb205-RBBP5-3xFlag: 3333, 1698, 1657 bp

pGreen-JMJD3-3xFlag: 1896, 1686, 1466, 866, 76 bp

pStb205-EZH2-3xFlag: 3333, 1675, 1210, 1112 bp

Figure 6.10 Cloning epigenetic editing genes into pKan^R-3xFLAG.
A. Plasmid maps showing Kan^R subcloning vectors including a gene of interest, **(i)** *Rbbp5*, **(ii)** *Jmjd3* or **(iii)** *Ezh2*, a 3xFLAG tag and additional restriction sites.
B. Gel image showing restriction digestion with *Bsm*BI of Kan^R subcloning vectors pStb205-RBBP5-3xFLAG, pGreen-JMJD3-3xFLAG and pStb205-EZH2-3xFLAG (lane 2, 3 and 4, respectively), confirming the gene cassette was successfully cloned. The expected restriction digestion pattern is shown.

6.2.1.5 Generation of *Nrp1* TALE fusion vectors

TALE fusions constructs were generated by Golden Gate cloning using a modified protocol and plasmids from the Voytas Lab (Cermak et al., 2011). The pTAL2 vector (Cermak et al., 2011) contained one single set of *Bsm*BI Golden Gate restriction sites flanking a *LacZ* gene and flanked by the TAL N- and C- terminal domains. This vector was designed to fuse TALE arrays and contains Gateway cloning sites flanking the TAL N- and C- terminal domains to allow subcloning of the TALE array into an expression vector.

To circumvent this final subcloning step, pTAL2 was modified to generate a TALE fusion expression vector, pTFE (Figure 6.11A) which would serve as a Golden Gate entry vector for inserts containing a promoter, TALE array and fusion domain of any design. The destination vector, pTFE, also contained a CMV-mTagBFP (BFP) expression cassette and ampicillin resistance gene. The sequences adjacent to the *Bsm*BI restriction sites for each insertion-destination site were designed to generate unique overhangs (coloured boxes in Figure 6.11A). These unique overhangs in the pTFE destination vector were only compatible with the *Bsm*BI overhangs in the appropriate subcloning vector (Appendix B, pp. 249). This allows for the orderly assembly of a promoter, TALE array and tagged gene of interest in an expression vector in a single ligation reaction.

To test the TALE fusion system, a panel of fusion constructs including a TALE array targeting the bivalent promoter of *Nrp1* (tNrp1) were generated. In these constructs, tNrp1 was designed to fuse with a 3x FLAG tagged *Rbbp5*, *Jmjd3* and *Ezh2*, or a TALE-Only-A negative control (Figure 6.11B). The Nrp1-TALE arrays were designed to target the same *Nrp1* promoter sequence as a previously published set of TALEs, which were shown to efficiently target the *Nrp1* promoter (Therizols et al., 2014).

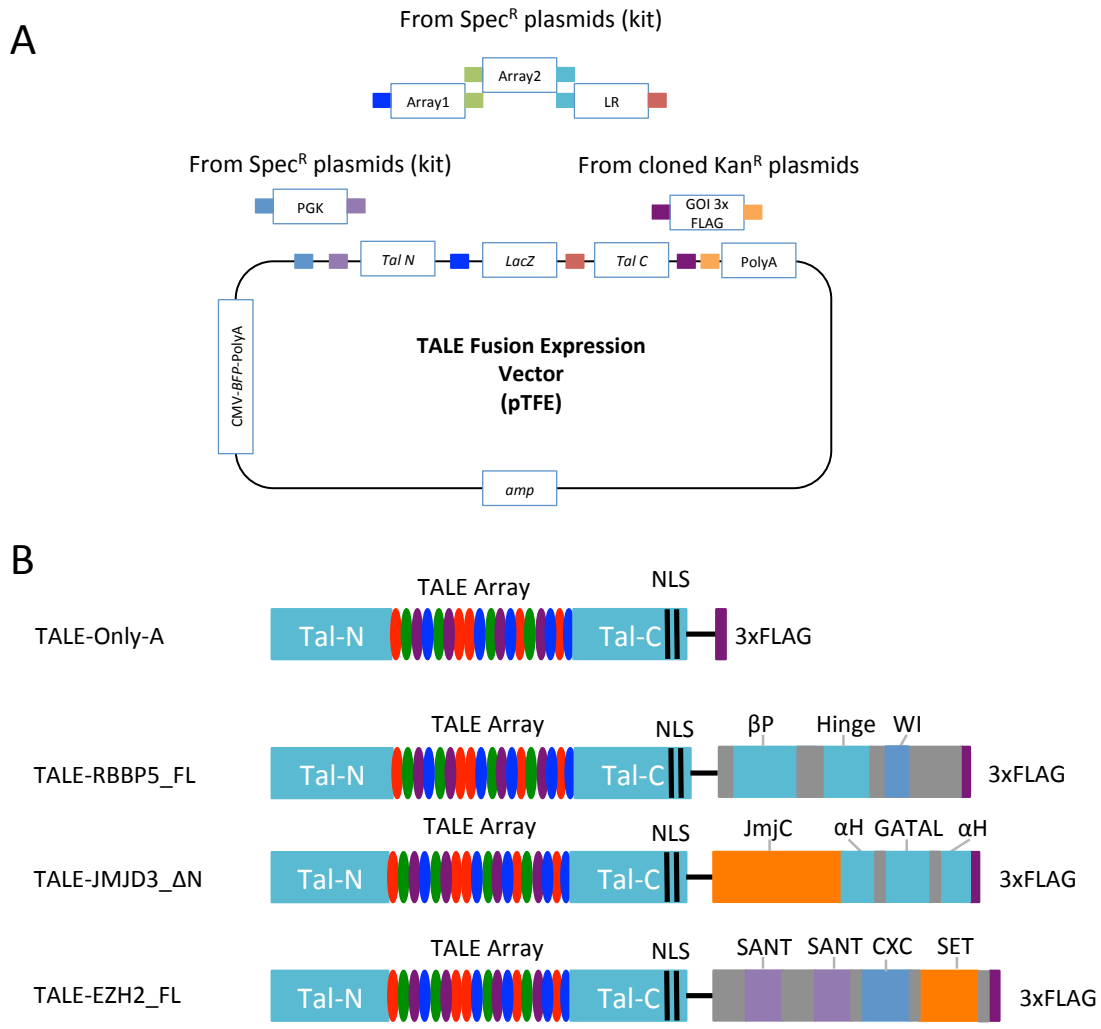


Figure 6.11 Generation of TALE fusions.
A. Cloning strategy for fusing promoter, TALE array and gene of interest (GOI) tagged with 3xFLAG into modified TALE Fusion Expression Vector, pTFE. All plasmids are digested with an excess of *Bsm*BI and then ligated at equimolar concentrations before transforming into chemically competent E-coli and expanding on ampicillin LB/agarose plates. The use of counterselective antibiotic resistant cassettes decreases the number of background colonies during cloning.
B. Schematic showing the TALE Array fusions generated in this study. TALE arrays, which were designed to target the *Nrp1* promoter, were fused to full-length cDNA of Rbbp5 (TALE-RBBP5_FL), the catalytic c-terminus of *Jmjd3* cDNA (TALE-JMJD3_ΔN) and the full-length cDNA of *Ezh2* (TALE-EZH2_FL). All of these fusions, including a TALE-Only negative control (TALE-Only-A) were tagged with a 3xFLAG tag. TAL-N, TAL N-terminus; TAL-C, TAL-C terminus; NLS, nuclear localisation signal; 3xFLAG, three times FLAG epitope tag; βP, beta propeller region; Hinge, WI, WDR5 interaction motif; JmjC, Jumonji-C domain (catalytic domain); αH, alpha helical domain; GATAL, a DNA binding domain which interacts with a G-A-T-A DNA motif; SANT, SANT domain for protein-histone interactions; CXC, a DNA binding domain which interacts with a C-X-C DNA motif; SET; domain involved in substrate recognition and catalytic activity.

Briefly, all the plasmids required in cloning the final TALE fusion constructs were digested with an excess of restriction enzyme *Bsm*BI and ligated to achieve the desired final construct. Ligated products were transformed into chemically competent *E. coli* and plated on LB/agar supplemented with ampicillin. This ensured that none of the subcloning vectors, which were either kanamycin or spectinomycin resistant, would be able to expand under ampicillin selection. Resulting colonies were screened by PCR (data not shown) and confirmed by restriction digestion with *Bsm*BI and Sanger sequencing.

Firstly, Golden Gate reactions were performed as described above to insert either PGK or the *Nrp1* TALE array (tNrp1) into pTFE to generate the plasmids pPGK-TFE and pTFE-tNrp1 (Figure 6.12). These three vectors could then be used in further Golden Gate reactions to generate the final fusion vectors (Figure 6.13), pPGK-tNrp1-GOI-3xFLAG with either *Rbbp5*, *Jmjd3*, *Ezh2,* or the TALE-Only-A negative control. The cloning efficiency of the 3xFlag tagged gene of interest was approximately 60%, however, the insertion efficiency of the PGK promoter was much lower at 1-6%, possibly due to the high GC content of the PGK promoter. The sequence of the final fusion constructs pPGK-tNrp1-RBBP5-3xFlag, pPGK-tNrp1-JMJD3-3xFlag, pPGK-tNrp1-EZH2-3xFlag and pPGK-tNrp1-3xFlag was confirmed by restriction digestion with *Bsm*BI (Figure 6.13B) and Sanger sequencing (data not shown).

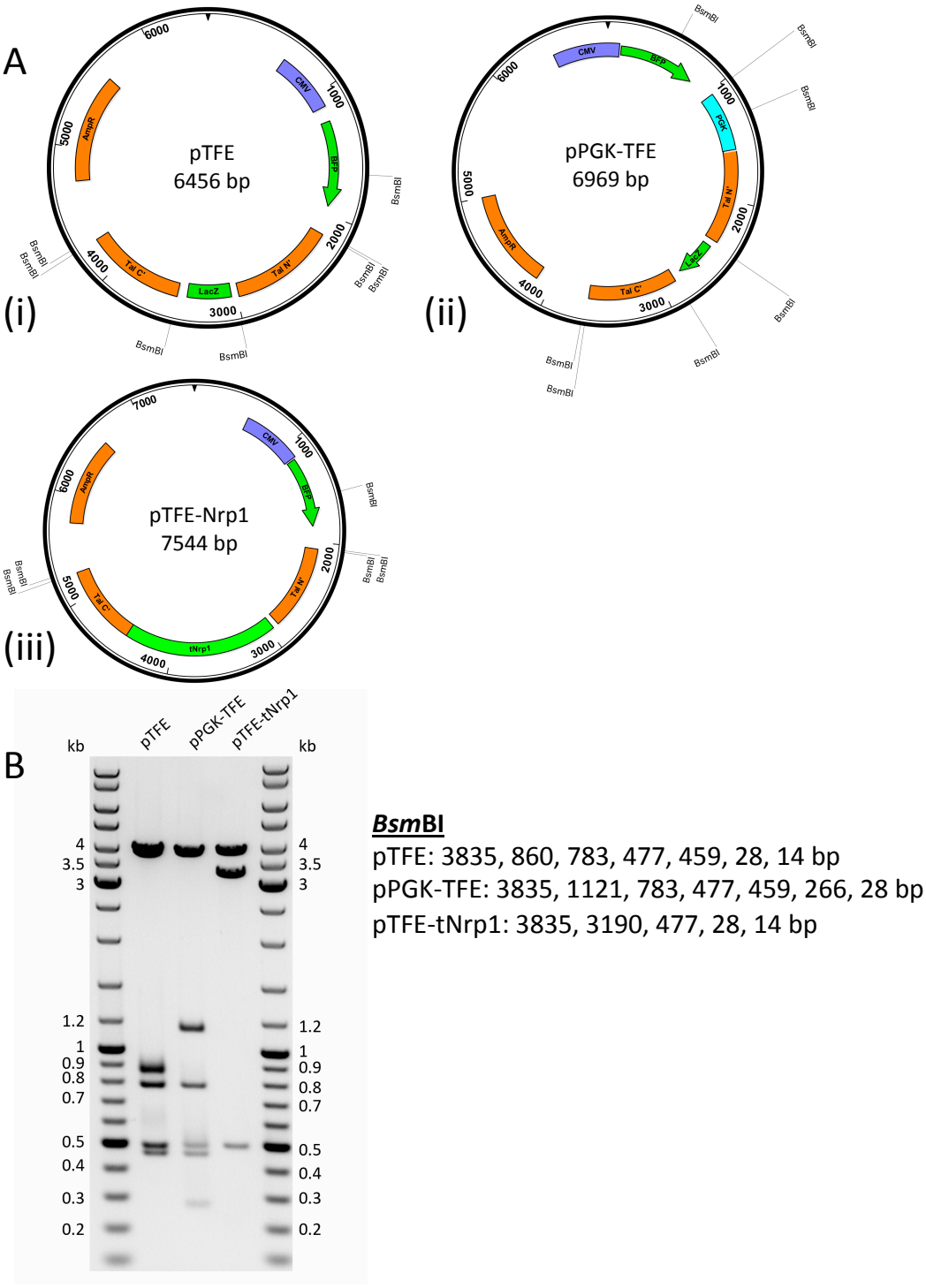
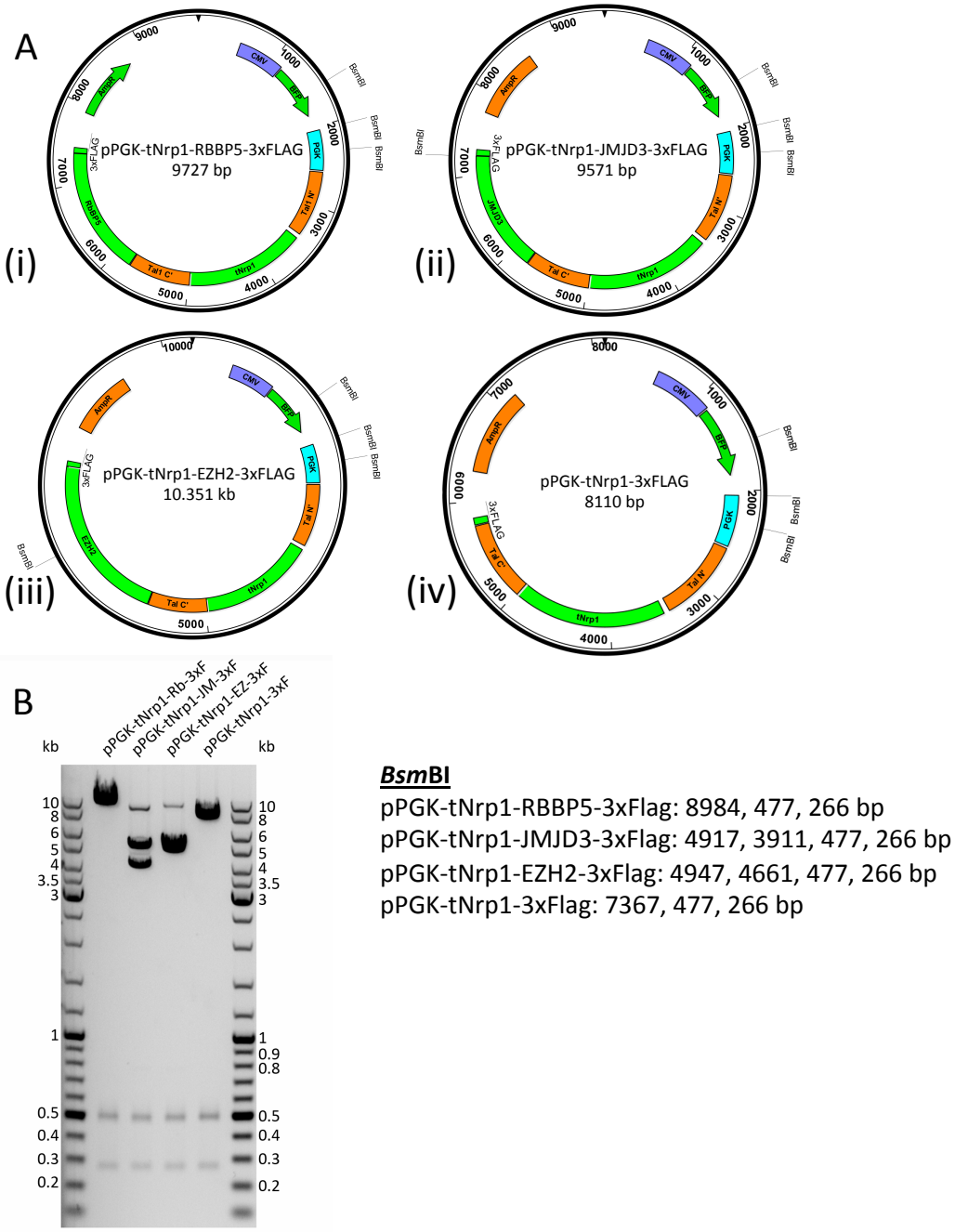


Figure 6.12 TALE fusion expression plasmid and subcloning vectors.
A. Plasmid maps showing **(i)** pTALE-Fusion-Expression (pTFE) and subcloning vectors **(ii)** pPGK-TFE and **(iii)** pTFE-tNrp1.
B. Gel image showing restriction digestion with *Bsm*BI of pTFE, pPGK-TFE and pTFE-tNrp1. The expected restriction digestion pattern is shown.



***Bsm*BI**

pPGK-tNrp1-RBBP5-3xFlag: 8984, 477, 266 bp

pPGK-tNrp1-JMJD3-3xFlag: 4917, 3911, 477, 266 bp

pPGK-tNrp1-EZH2-3xFlag: 4947, 4661, 477, 266 bp

pPGK-tNrp1-3xFlag: 7367, 477, 266 bp

Figure 6.13 Cloning final *Nrp1* TALE fusion plasmids

A. Plasmid maps showing final TALE fusion plasmids including the PGK promoter, tNrp1 and a gene of interest tagged with 3xFLAG are shown **(i-iii)**. A negative control TALE fusion containing the PGK promoter, tNrp1 tagged with 3xFLAG (also here called TALE-Only-A) is also shown **(iv)**.

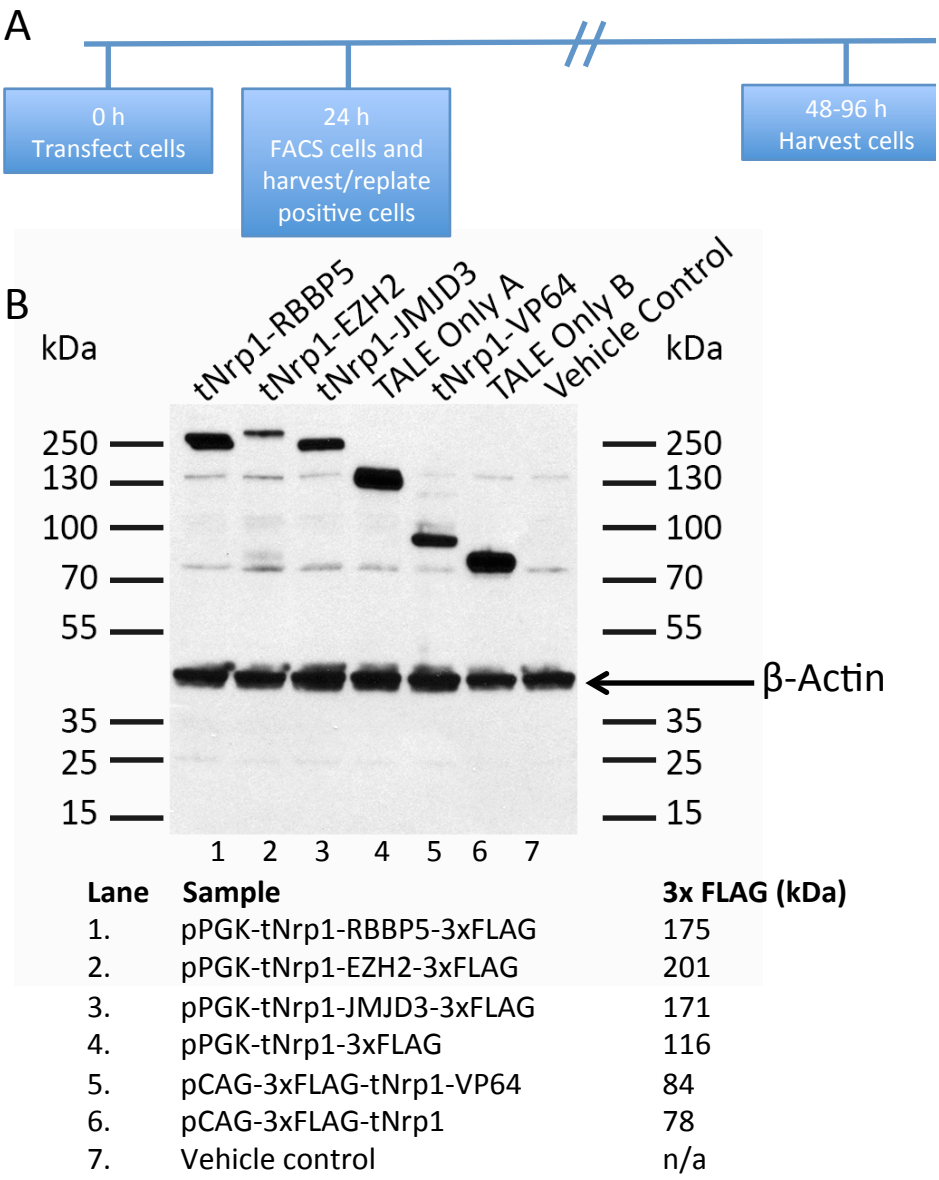
B. Gel image showing restriction digestion with *Bsm*BI of final TALE fusion plasmids confirming successful cloning. The expected restriction digestion pattern is shown. An unexpected band of >8 kb was detected in digestions of final fusion plasmids containing *Jmjd3* and *Ezh2*, possibly due to incomplete plasmid digestion, but these clones were confirmed by Sanger sequencing and determined to be appropriate for use in future experiments.

6.2.2 Effects of tNrp1 TALE fusions on *Nrp1* gene expression

6.2.2.1 Transfection strategy and validation of construct expression

In this section, I describe the strategy used to test the ability of tNrp1 TALE fusions to modify *Nrp1* expression and histone methylation marks *in vitro* (Figure 6.14A). Firstly, the fusion constructs were transfected into WTH3 mES cells and positively transfected cells were isolated 24 h after transfection by FACS. A time course experiment was performed during which positively transfected cells were harvested for analysis 24-96 h after transfection. The four tNrp1 fusion constructs generated in this thesis (Figure 6.13A) were tested using this strategy alongside two published control tNrp1 fusions (Therizols et al., 2014). The tNrp1 fusions included tNrp1-VP64 (a TALE targeting the Nrp1 promoter with VP64) as a positive control and tNrp1-Δ (also here called TALE-Only-B), as a negative control. These control fusions contain a 2A-eGFP cassette and also contain a 3xFLAG tag.

To confirm the expression of TALE fusion proteins, WTH3 mES cells were lipofected with each of the TALE fusion contructs, together with a negative vehicle control (Lipofectamine reagent mixed with OptiMEM and TE only). The bulk, unsorted population was harvested 24 h later and lysed for protein analysis. A western blot was performed with the cell lysates using a FLAG antibody. A β-actin antibody was used as a loading control (Figure 6.14B). The western blot confirmed all fusion constructs were expressed in the cells 24 h after transfection (Figure 6.14B).



Lane	Sample	3x FLAG (kDa)
1.	pPGK-tNrp1-RBBP5-3xFLAG	175
2.	pPGK-tNrp1-EZH2-3xFLAG	201
3.	pPGK-tNrp1-JMJD3-3xFLAG	171
4.	pPGK-tNrp1-3xFLAG	116
5.	pCAG-3xFLAG-tNrp1-VP64	84
6.	pCAG-3xFLAG-tNrp1	78
7.	Vehicle control	n/a

Figure 6.14 Experimental strategy and TALE fusion construct expression.
A. Experimental strategy for TALE fusion transfection experiments. WTH3 humanised mES cells were transfected with TALE fusion plasmids (Time: 0 h). Approximately 24 h later, transfected cells were harvested. Positive cells were isolated by FACS and either processed for downstream analysis (RNA, ChIP) or maintained in culture until a later time point. Cells maintained in culture were harvested and processed for downstream analysis at approximately 24 h time points up to 96 h post-transfection.
B. WTH3 cells transfected with TALE fusion vectors and a vehicle control were harvested (unsorted) 24 h after transfection. A western blot was performed on cell lysates using a FLAG antibody confirming construct expression. Expected protein sizes of TALE fusions are shown on the right. A β-actin loading control showed similar protein loading in all lanes (expected size: 40 kDa).

For RNA and chromatin analysis, 4-7 wells in 6 well plates were lipofected with each plasmid sample and only positively transfected cells (isolated by FACS) were analysed in an effort to amplify any detected change in signal due to expression of the TALE fusions. First, live cells were gated in P1 (Figure 6.15) and either BFP or GFP positive cells were isolated (Figure 6.15A and B) from the unstained cell populations (Figure 6.15C). Overall, transfection efficiencies of plasmids containing the BFP marker were lower (<20% of live cells) than those containing the GFP marker (20-70% of live cells) (Figure 6.16A). Notably, the cell death following FACS was extremely high (~65-95%) in all experiments (Figure 6.16B). Following FACS in independent biological replicates 1-3, the total number of positively transfected cells in each sample was split between a number of different wells for harvesting at different time points. In an attempt to reduce cell death, in replicate 4 and 5 the same number of positively transfected cells for each sample was plated at a higher density in 96 or 24 well plates. However, cell recovery still remained low in replicate 4 and 5 (Figure 6.16B).

Following FACS, only cells positive for fluorescence were re-plated for harvesting at later time points. Despite being positive for fluorescence at the time of FACS, mES cells in all samples began losing fluorescence even 24 h after FACS (Figure 6.17). At 96 h after FACS very few fluorescent mES cells can be detected by fluorescent microscopy in positively transfected samples (Figure 6.18). In a similar transfection experiment of Plat-E cells (a modified HEK293T cell line) using the same plasmids, expression of the fluorescent proteins was maintained at very high levels in the majority of cells up to 96 h after transfection (data not shown).

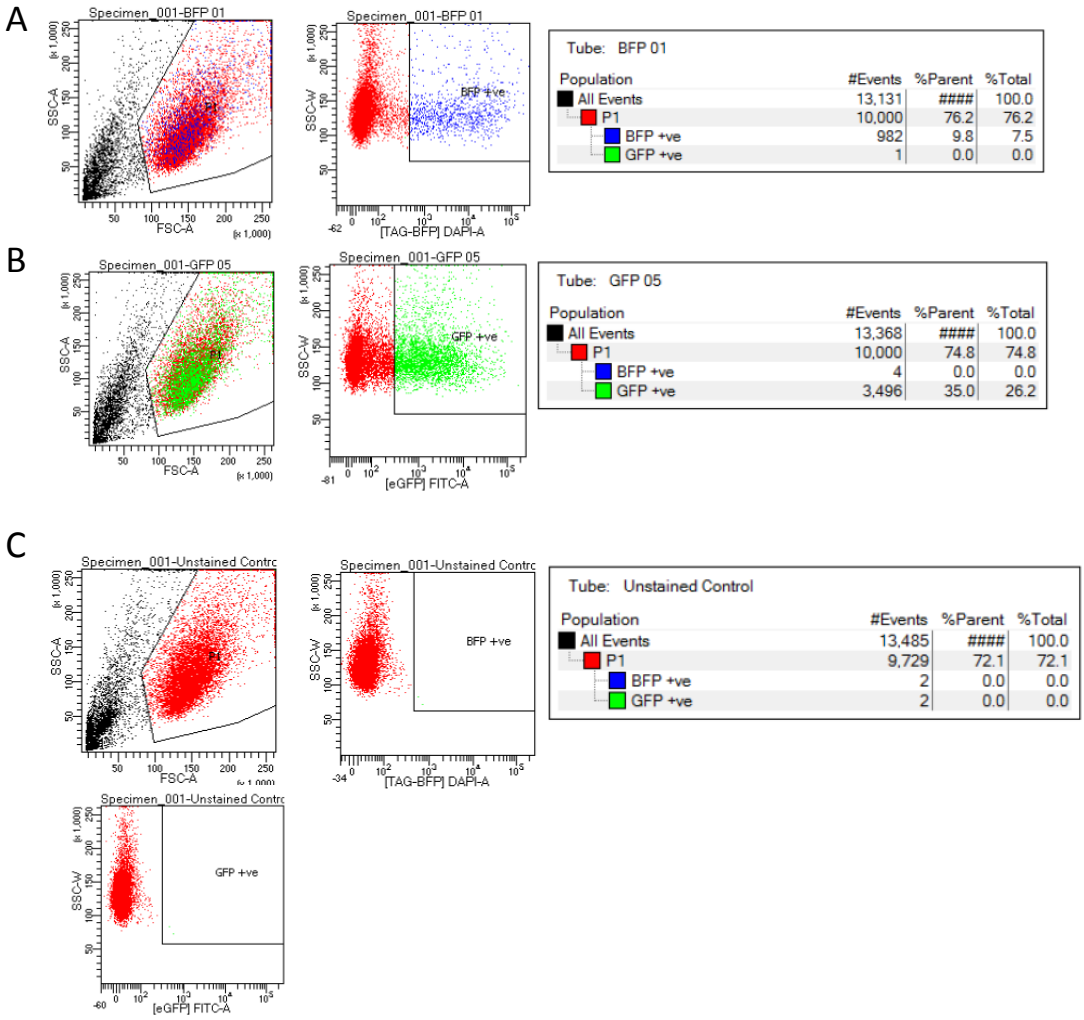


Figure 6.15 Gating strategy in TALE fusion experiments.
Representative examples of the gating strategy used during FACS are shown here. In all cases, live cells were gated excluding doublets and cell debris (P1). Cells positively expressing BFP **(A)** or GFP **(B)** were collected and used for downstream analysis. Cells transfected with a vehicle control showed no fluorescence **(C)**.

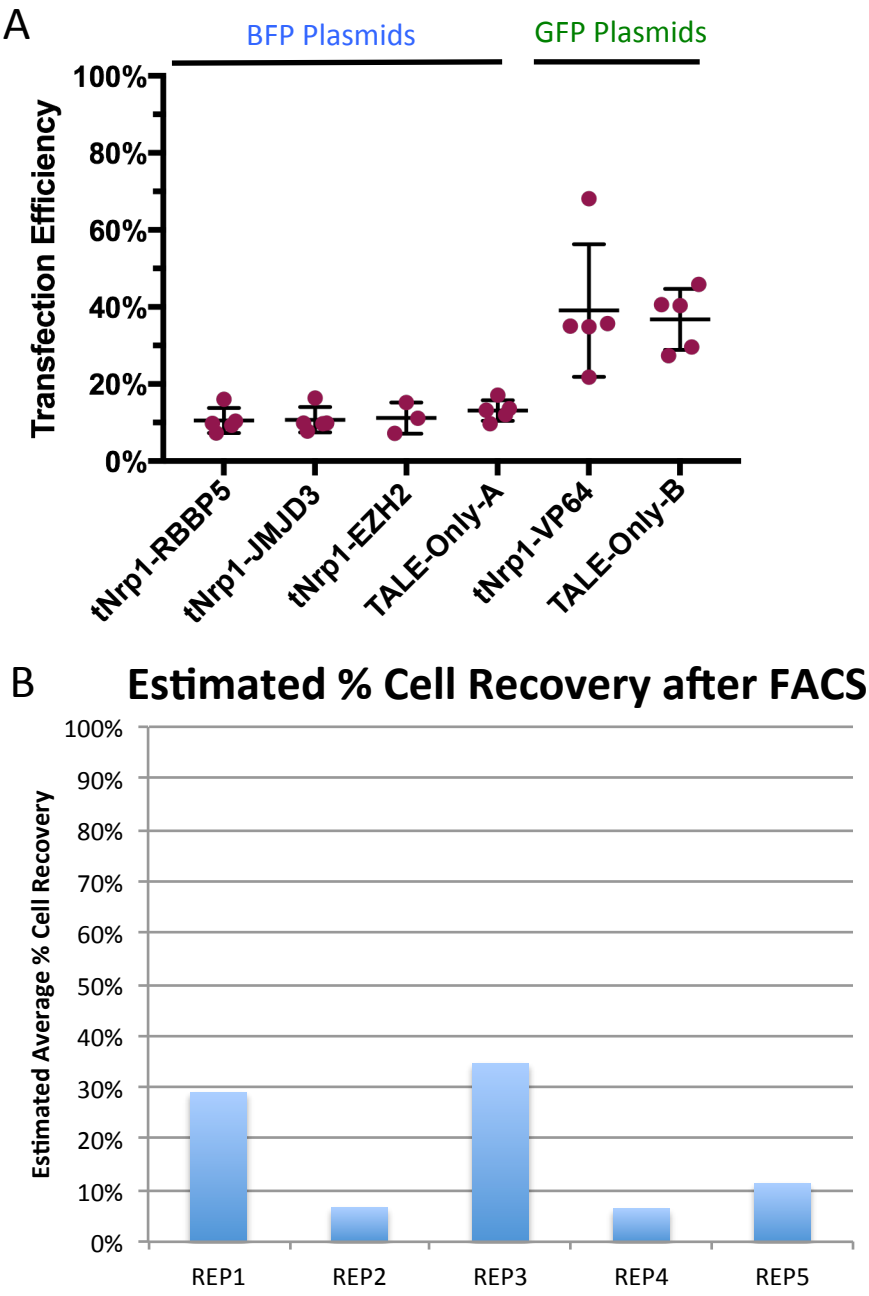


Figure 6.16 Transfection efficiency and cell death in TALE fusion experiments
A. Graph plotting transfection efficiencies of TALE fusion plasmids in independent biological replicates as determined by FACS data. TALE fusion plasmids containing a BFP marker transfected at lower efficiencies compared to those containing a GFP marker. However, there was a higher degree of variance between transfection efficiencies in biological replicates of plasmids containing a GFP marker vs. those containing a BFP marker.
B. Graph plotting the average estimated cell recovery 24 h after FACS in each biological replicate of transfection experiments. Cell recovery was estimated based on the number of cells collected by FACS, the number of cells harvested at a time point at least ~24 h following FACS and the average time of mES cell duplication. Cell death is >60% in each biological replicate.

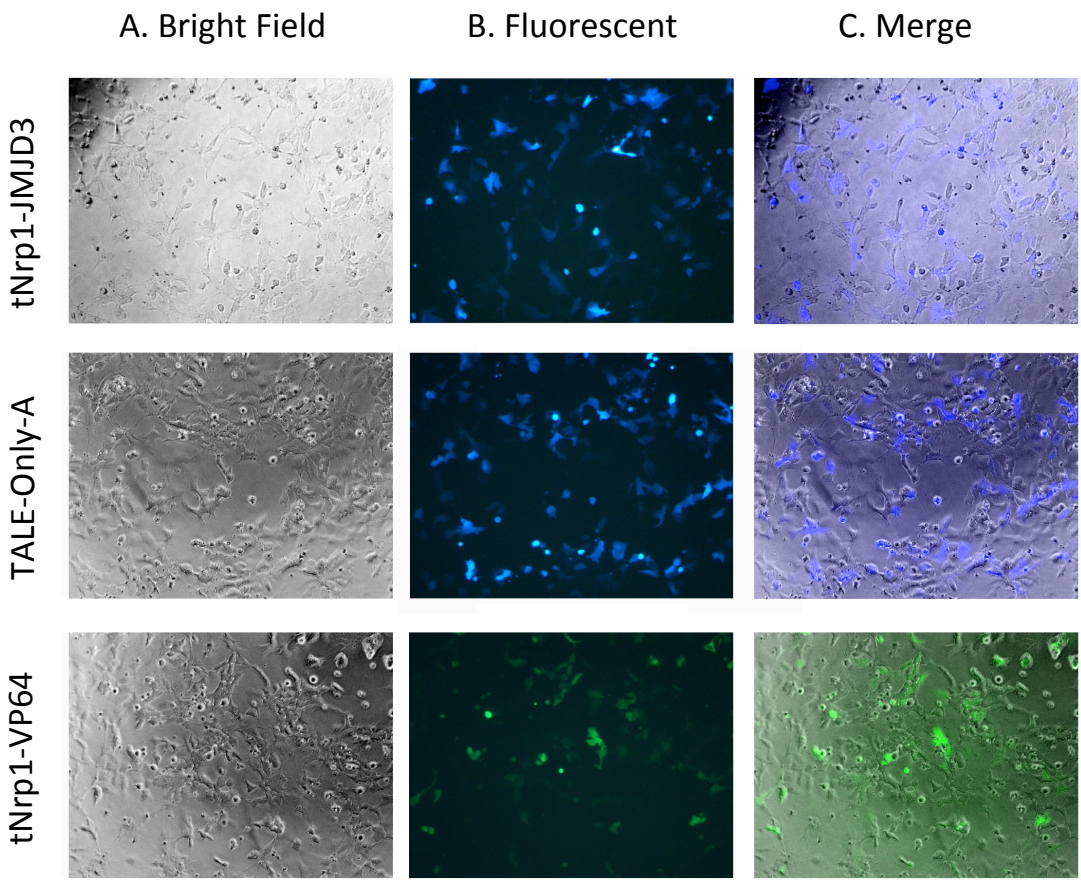


Figure 6.17 Fluorescent microscopy images of tNrp1 fusion plasmid transfected cells at 48 h post transfection.
Representative images **(A-C)** from tNrp1 fusion plasmid transfection experiments showing transfected cells from replicate 5 in 96 well plates 24 h following FACS. Fluorescent images **(B-C)** show that only a proportion of cells maintain fluorescence after FACS. The images were taken using a Zeiss axiovert 25s inverted microscope at 10x magnification using a DAPI filter to detect BFP fluorescence and a FITC filter to detect GFP fluorescence. The merge images were generated using the ImageJ software package.

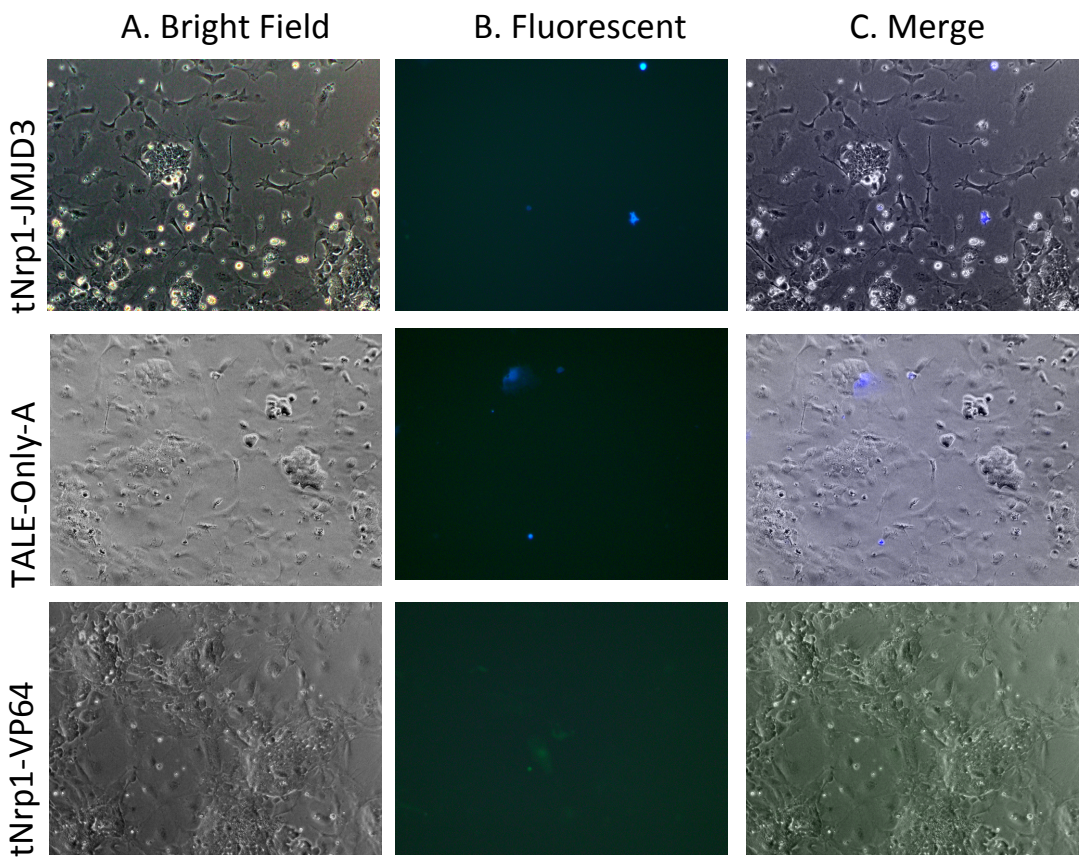


Figure 6.18 Fluorescent microscopy images of tNrp1 fusion plasmid transfected cells at 96 h post transfection.
Representative images **(A-C)** from tNrp1 fusion plasmid transfection experiments showing transfected cells from replicate 4 in 24 well plates 96 h following transfection. Fluorescent images **(B-C)** show that the majority of cells completely lose fluorescence 96 h after transfection. The images were taken using a Zeiss axiovert 25s inverted microscope at 10x magnification using a DAPI filter to detect BFP fluorescence and a FITC filter to detect GFP fluorescence. The merge images were generated using the ImageJ software package.

6.2.2.2 Monitoring changes in Nrp1 expression following the transfection of vectors containing TALE fusions

WTH3 mES cells lipofected with TALE fusion plasmids were harvested between 24 and 96 h after transfection. Although reporter expression was low at 96 h post transfection, the time course was extended to this point for certain vectors to see if there was a latent effect of TALE fusion expression on *Nrp1* transcription. RNA was isolated from harvested samples and reverse transcribed to produce cDNA templates that were then analysed by qPCR (Taqman). The relative expression of *Nrp1*/*Gapdh* for each transfection was normalised to the mean of the vehicle control at each time point (Figure 6.19).

Nrp1 expression was at least ten times higher in the tNrp1-VP64 fusion transfected samples relative to the TALE-only B control at both 24 and 48 h post-transfection (Figure 6.19A, Figure 6.20A). However, *Nrp1* expression was highly variable in samples transfected with tNrp1-VP64 at 72 and 96 h (Figure 6.19A).

Interestingly, *Nrp1* expression was moderately higher in cells transfected with the tNrp1-JMJD3 fusion relative to the negative TALE-Only-A control at 96 h post- transfection (Figure 6.19B, Figure 6.20B). *Nrp1* expression was similar in cells transfected with the tNrp1-EZH2 and -RBBP5 fusions relative to the TALE-Only A control at every time point analysed (Figure 6.19B).

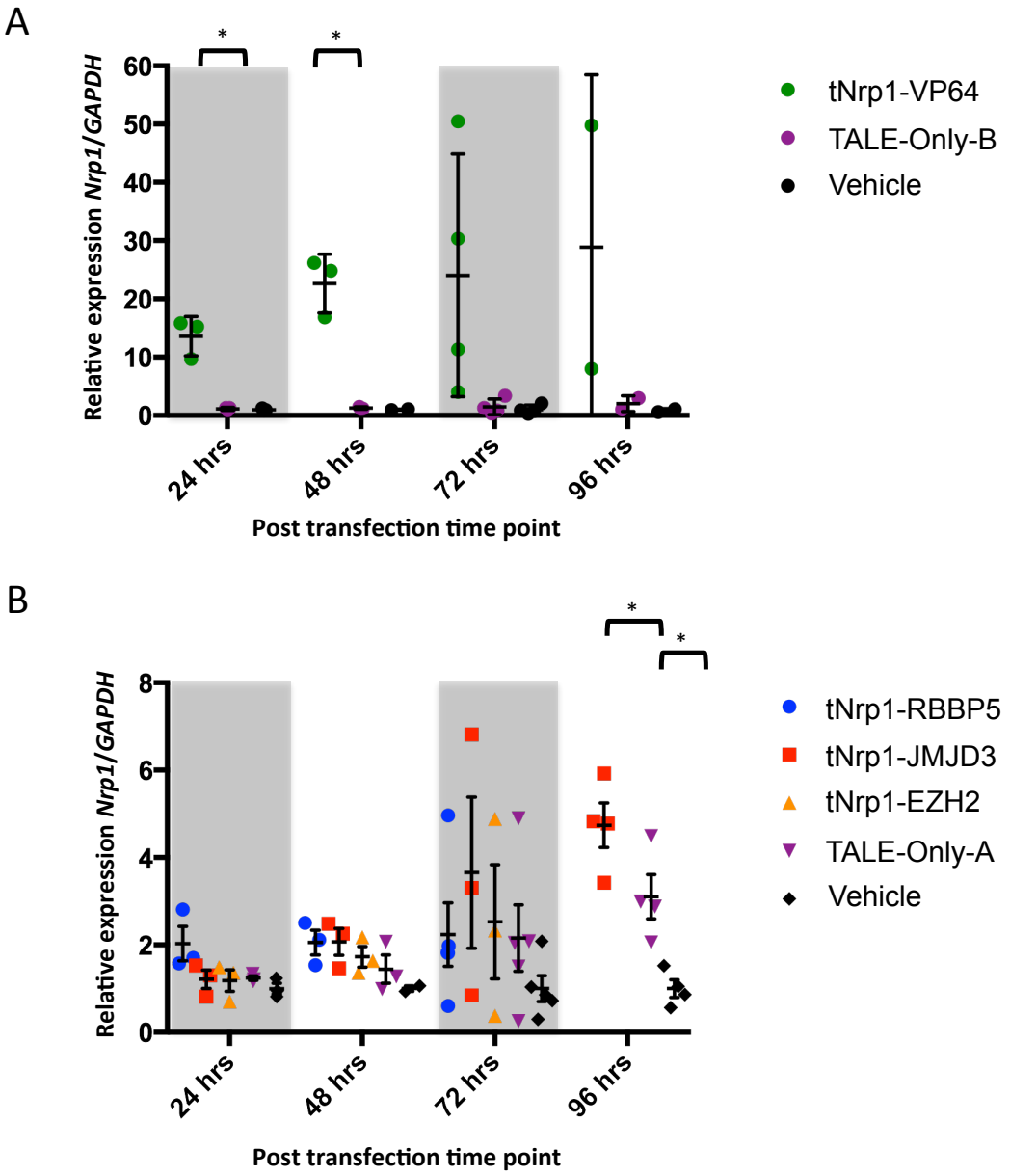


Figure 6.19 RT-qPCR analysis of *Nrp1* expression in transfected cells.
Dot plots A and B depict the changes in relative enrichment of *Nrp1*/*Gapdh* at 24-96 h post-transfection for each of the TALE fusion transfection samples. Each point represents expression changes from each of the 2-5 paired independent biological replicates performed for each transfection. A '*' indicates a statistically significant difference ($p > 0.05$) was observed between two samples.
A. The relative enrichment of *Nrp1*/*Gapdh* in the positive control tNrp1-VP64, negative control TALE-Only-B and vehicle control samples. The error bars represent standard deviation from the mean.
B. The relative enrichment of *Nrp1*/*Gapdh* in samples tNrp1-RBBP5, -JMJD3 and –EZH2, and the negative control TALE-Only-A and vehicle control samples. The error bars represent standard error of the mean.

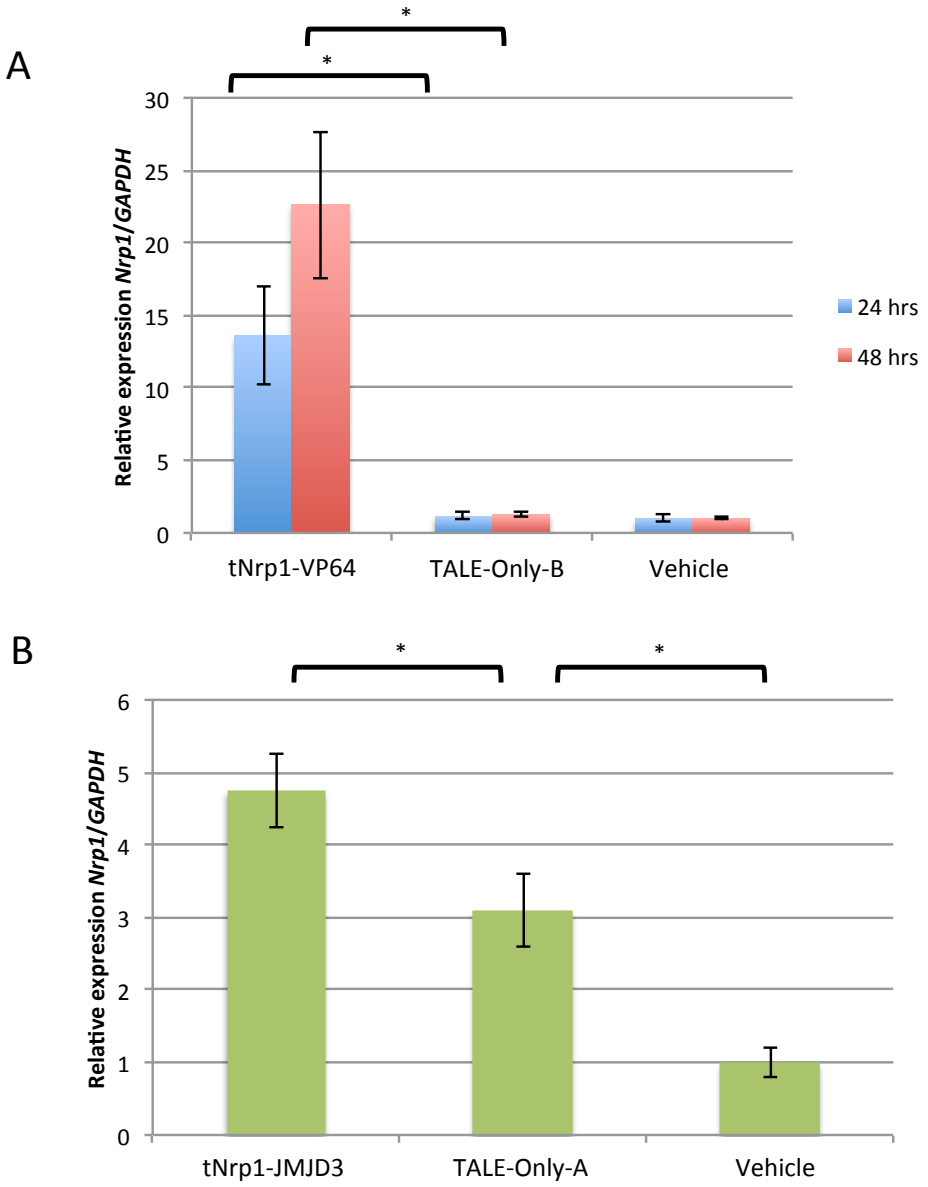


Figure 6.20 *Nrp1* expression in cells transfected with JMJD3 and VP64 tNrp1 fusions.
A. Bar chart plotting RT-qPCR data showing *Nrp1* expression relative to *Gapdh* at 24 and 48 h post transfection. Error bars indicate standard deviation over three independent biological replicates. A '*' indicates a statistically significant difference was observed between two samples using a paired two tailed students T-test. *Nrp1* expression is elevated in cells transfected with the VP64-tNrp1 fusion relative to the TALE-Only-B negative control.
B. Bar chart plotting RT-qPCR data showing *Nrp1* expression relative to *Gapdh* at 96 h post transfection. Error bars indicate standard error of the mean over four independent biological replicates. *Nrp1* expression is elevated in cells transfected with the JMJD3-tNrp1 fusion relative to the TALE-Only-A negative control. A '*' indicates a statistically significant difference was observed between two samples using a paired two tailed students T-test.

6.2.3 Effects of tNrp1 TALE fusions on bivalent histone modifications at Nrp1 promoter

As statistically significant increases in *Nrp1* gene expression were detected in samples transfected with tNrp1-VP64 at 24 h and tNrp1-JMJD3 at 96 h, ChIP-qPCR analysis was performed on these samples, and the appropriate TALE-Only negative controls, at these time points. Three independent biological replicates of lipofection, FACS and harvesting of the tNrp1-VP64 and tNrp1-JMJD3 constructs and the TALE-Only-A and -B controls were performed, each on 6 wells of WTH3 mES cells. The ChIP for each independent biological replicate was also performed on independent days. The ChIP was performed on all samples using antibodies against the histone modifications, H3K4me3 and H3K27me3. ChIP-qPCR (SYBR Green) amplicons were designed to target a number of positions along the *Nrp1* genomic locus (shown as orange arrows in Figure 6.3A). The expected results, based on analysis of ChIP-seq data from the "Bruce-4" mES cell line on the UCSC genome browser (Figure 6.3A), are listed in Table 6.2.

Table 6.2 Expected results from ChIP qPCRs using antibodies for bivalent histone modifications, H3K4me3 and H3K27me3. The locations of the Nrp1 amplicons are indicated as numbered orange arrows in the UCSC ChIP seq data plot in Figure 6.3A. Primer set 1 is 4.8 kb upstream of the *Nrp1* start codon and set 2 is 98 bp upstream from the tNrp1 binding site at the *Nrp1* promoter. Expected results are based on published ChIP-seq data (UCSC). Negative control amplicons (5` mouse α-globin, 3` mouse α-globin and *Gapdh*) for both antibodies were also used to estimate background signal.

Modification	1. Nrp1 UpA	2. Nrp1 UpB	3. Nrp1 ex2	4. Nrp1 ex4	β-Actin CpG	Gata6
H3K4me3	None	Medium	None	None	High	Medium
H3K27me3	Low	Medium	Low	Low	None	Medium

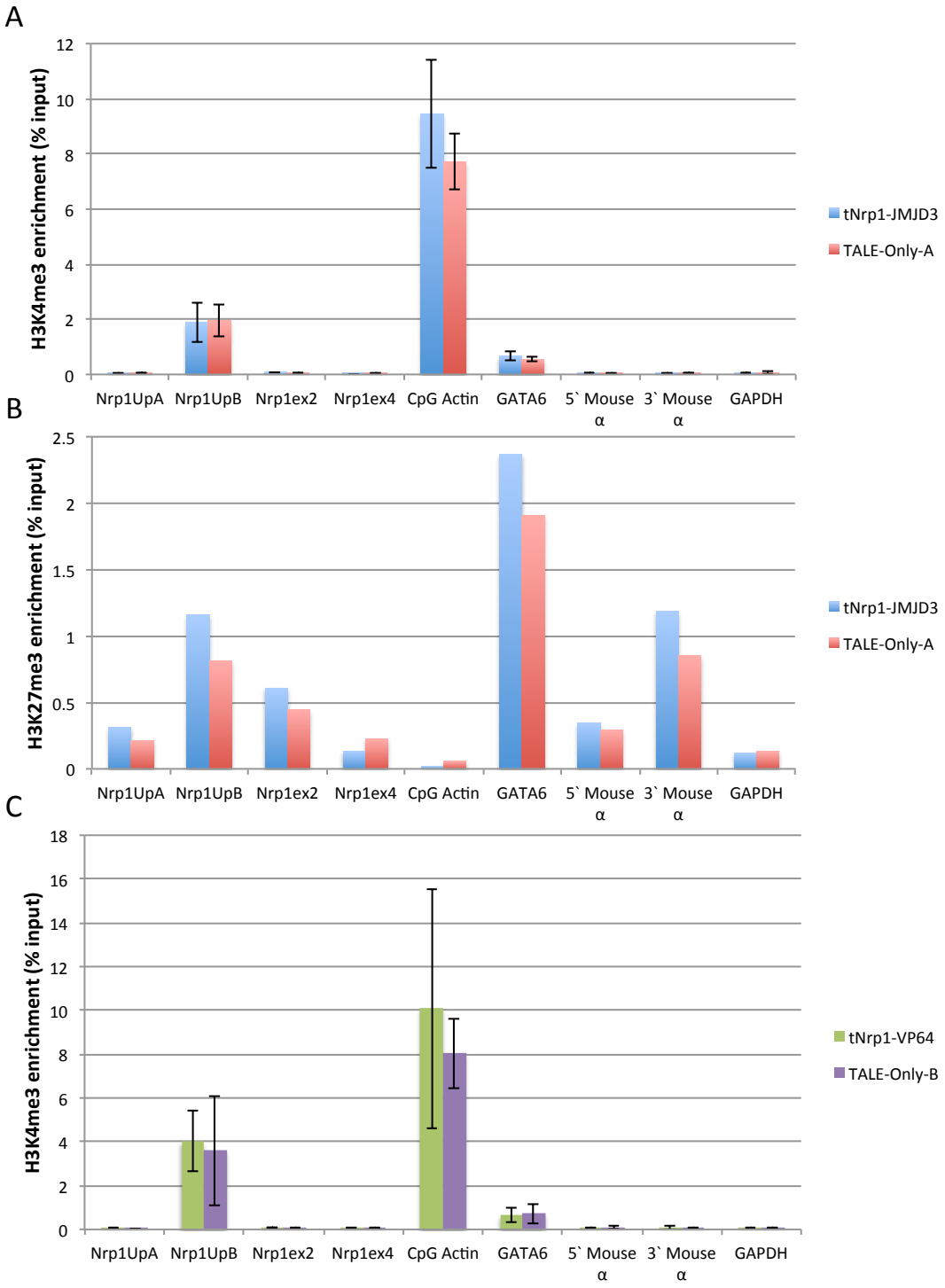


Figure 6.21 ChIP-qPCR Data from tNrp1 fusion transfections.
A. Graph plotting H3K4me3 enrichment as a percentage of input chromatin for 3 independent biological replicates did not show any significant difference increase between H3K4me3 levels in JMJD3-tNrp1 vs. TALE-Only-A samples harvested at 96 h post-transfection.
Figure legend is continued on the next page.

The TALE fusion tNrp1-JMJD3 was expected to reduce the levels of H3K27me3 at the *Nrp1* promoter (Nrp1 UpB) and the increase in expression of *Nrp1* suggested that H3K4me3 levels may be increased at the promoter. However, no change in H3K4me3 at the *Nrp1* promoter was detected in tNrp1-JMJD3 treated cells relative to the TALE-Only-A negative control (Figure 6.21A). ChIP-qPCR analysis was performed on all samples from the H3K27me3 ChIP on tNrp1-JMJD3 and TALE-Only-A control samples, however, qPCR amplification was only detected from the IP samples from one out of the three biological replicates although amplification from the input samples was comparable to other experiments. In this sample there is no detectable decrease in H3K27me3 at the *Nrp1* promoter in the tNrp1-JMJD3 treated sample vs. the TALE-Only-A Control (Figure 6.21B). However, it is important to note that detected H3K27me3 levels in the tNrp1-JMJD3 sample are higher at all the amplicons tested compared to the TALE-Only-A control (Figure 6.21B). No significant change was detected in H3K4me3 levels between the tNrp1-VP64 and TALE-Only-B samples (Figure 6.21C).

Figure 6.21 *Legend continued:*
B. Graph plotting H3K27me3 enrichment as a percentage of input chromatin for one biological replicate did not show any reduction of H3K27me3 levels in JMJD3-tNrp1 vs. TALE-Only-A samples harvested at 96 h post-transfection.
C. Graph plotting H3K4me3 enrichment as a percentage of input chromatin for 2 independent biological replicates did not show any significant difference between H3K4me3 levels in VP64-tNrp1 vs. TALE-Only-B samples harvested at 24 h post-transfection.

6.3 Discussion

The aim of this chapter was to develop an experimental strategy aimed at directly inferring the effects of modifying the identity of the chemical groups associated with histone tails. This strategy was developed using a model gene prior to interrogating the gene of interest in this thesis, *HBA2*, in order to identify areas in which the experimental design could be improved. Several positive conclusions have been reached over the course of this study. Firstly, an efficient cloning system for generating TALE fusions was developed and validated. This cloning system facilitates the rapid cloning of several modules including TALE arrays, promoters and enzymatic domains of interest, among other components. This system, with some modifications, will form the backbone of further studies. Secondly, I confirmed several published results, such as the induction of *Nrp1* by the Nrp1-VP64 fusion protein, and extended them by showing that the levels of H3K4me3 do not change after transfection of the plasmid harbouring this fusion gene. Thirdly, the expression of *Nrp1* was elevated in cells transfected with the tNrp1-JMJD3 fusion construct, which provides motivation for further investigation using the strategy outlined. Due to the initial experimental design, which included numerous positive and negative controls, the results obtained have also highlighted areas in which in the experimental strategy could be improved. Below, I will discuss the outcomes of these experiments and consider ways in which to optimise future experiments.

A cloning system for construction of TALE fusion expression constructs

Here, I developed a cloning system for the modular assembly of TALE fusion expression constructs, similar to MoClo (Weber et al., 2011) and based on a freely available Golden Gate TALE assembly kit (Cermak et al., 2011). Using this strategy, promoters of choice and different GOIs may be cloned into separate kanamycin resistant vectors. The plasmids are designed such that once the GOIs are cloned into the appropriate vector, they are fused at their 3` ends to a 3xFLAG tag, which can be used to reliably detect their expression *via* immunoblotting with a commercially available antibody (Figure 6.14B). TALE array entry plasmids can be assembled using the protocols outlined in Cermak *et al.* (2011). These components can then be inserted in a predetermined manner into the destination vector, pTFE, by digesting

them with an excess of the type IIS restriction enzyme (Figure 6.2B). I demonstrated the efficiency and usefulness of this system, which represents a key development of this study and will form the basis of future experimental strategies.

Transfection efficiency of tNrp1 fusions

The fusion proteins generated in this study were designed to target the promoter of *Nrp1*, a bivalent gene in mES cells, using a validated TALE sequence (Therizols et al., 2014). The activator tNrp1-VP64 and negative control tNrp1-Δ (Therizols et al., 2014) were used as controls to show that transient transfection of the targeted transactivator into WTH3 mES cells could efficiently induce *Nrp1* expression. The expression of the tNrp1 fusion proteins in WTH3 mES cells following transfection was confirmed by western blot (Figure 6.14B).

Unexpectedly, the transfection efficiency (as estimated by FACS) of the TALE fusion constructs generated in this chapter was much lower (~10%) than the transfection efficiency of the control TALE fusions from the Bickmore lab (~40%; Figure 6.16). The TALE fusion constructs from the Bickmore lab were marginally smaller (7.3-7.5 kb) than those generated in this chapter (8.1-10.4 kb) which may have had an effect on transfection efficiency. The fluorescent marker used in the constructs generated in the chapter, mTagBFP (Subach et al., 2008), is comparably bright relative to eGFP, the fluorescent marker used in the constructs from the Bickmore lab, so one would expect them both to be detected by FACS with the same efficiency (Kremers et al., 2011). However, mTagBFP is somewhat less photostable than eGFP (Kremers et al., 2011). It is also important to consider that the *eGFP* gene is driven by a CAG promoter in the Bickmore constructs while the *mTagBFP* gene in the constructs generated here is driven by a CMV promoter. The CAG promoter drives constitutive and consistently high expression in a number of cell types while CMV promoter driven expression can vary widely between different cell types (Qin et al., 2010). This may have led to a lower detection rate by FACS of transfection efficiency and the use of another promoter may drive higher reporter expression. Another method that could have been used to compare transfection efficiency of these plasmids independent of fluorescence would have been to normalise the gene expression to a region homologous to all the transfected plasmids, such as a region

within the TAL N- and C- termini, however this approach was not pursued in this study.

After transfection, the cells that were actively expressing a fluorescent marker were sorted by FACS and returned to culture. Fluorescence in the transfected mES cells was extremely rare just four days after transfection (Figure 6.18). However, Plat-E cells that were transfected with the same constructs maintained fluorescence in a high proportion of cells up to 96 h after transfection (data not shown). The Plat-E cell line, a modified HEK293T cell line, expresses the Simian virus 40 (SV40) large T-antigen which allows for the replication and maintenance of plasmids containing a SV40 origin of replication (Dubridge et al., 1987; Morita et al., 2000). However, none of the plasmids used in this study contain a SV40 origin of replication. It is thus unclear why the expression from these plasmids was maintained in the Plat-E cells but completely disappeared in the WTH3 mES cells 96 h after transfection.

Further improvements of the experimental strategy

This study has formed the basis to improve and optimise this experimental strategy. Methods to improve the low transfection efficiency observed in these experiments are discussed in the above section. The high degree of variability observed in the expression of *Nrp1* following transfection severely limited the ability to observe differences between different treatment conditions (Figure 6.19). A number of strategies could be attempted to limit this variability, including use of 2i media for mES cell culture, removal of the FACS step and generation of stably transfected cell lines.

It is well known that 2i media can reduce the variability of expression of genes involved in differentiation in mES cell lines (Guo et al., 2016; Ying et al., 2008). Two inhibitors, of the FGF receptor tyrosine kinases and the ERK cascade, are used in 2i media to enforce a ground state of pluripotency in ES cells (Ying et al., 2008). Although it has been reported that mES cells cultured in 2i media are harder to transfect than those cultured in normal mouse ES media (Tamm et al., 2013), recently optimised methods allow highly efficient transfection efficiency of mES cells cultured in 2i media (Tamm et al., 2016). Furthermore, the use of 2i medium

has been shown to reduce the number of bivalent genes observed in mES cell culture (Guo et al., 2016), possibly resolving any observed bivalency that is strictly due to a epigenetically heterogeneous cell population rather than true bivalency. The use of 2i medium could reduce the rate of heterogeneous *Nrp1* expression or histone methylation, providing a cleaner background for the study of expression and histone methylation changes at *Nrp1*.

The use of a different transfection method could also be investigated with the view to boost TALE fusion expression rates and duration. The current system I have tested in this chapter uses transient transfection to introduce the TALE fusion constructs to mES cells, but expression of the fluorescent reporter in these constructs is undetectable 96 h after transfection (Figure 6.18). A stable transfection would allow sustained expression of the TALE fusion which could be combined with an inducible system to control TALE fusion induction in a temporal manner (Cano-Rodriguez et al., 2016; Kim and Eberwine, 2010). Similarly, the use of multiple TALE fusions to 'tile' the *Nrp1* promoter could have also led to a higher probability of affecting the histone methylation status and *Nrp1* expression rates (Maeder et al., 2013b; Mali et al., 2013a; Perez-Pinera et al., 2013). Indeed, the CRISPR/Cas9 system could be used as an alternative to quickly generate many sgRNAs that target a gene promoter to attract multiple copies of a dCas9 fusion protein (Hilton et al., 2015).

Extremely low rates of cell recovery were observed following FACS of transfected WTH3 mES cells in this chapter (Figure 6.16B) and similarly high rates of apoptosis following FACS of mES cells has been reported previously (Fukuda et al., 2006). High rates of cell apoptosis may have negative effects on the health of surviving cells and induce changes in gene expression or spontaneous differentiation. To avoid the necessity of FACS in expression studies following transfection, the expression values of the target gene could be normalised to a sequence homologous to all transfected constructs, as well as a housekeeping gene.

Effects of tNrp1 fusion on gene expression and histone modifications

The effects of expressing tNrp1 fused to the enzymatic domains of RBBP5, EZH2 and JMJD3 was tested by qRT-PCR and ChIP-qPCR. The expression of *Nrp1* alone was increased 96 h post-transfection, relative to negative controls, only with the vector containing tNrp1-JMJD3. Notably, *Nrp1* expression was increased in samples transfected with TALE-Only-A (a TALE-Only construct generated in this chapter) and -B (a TALE-Only construct generated in Therizols *et al.* (2014)) relative to the vehicle control. This increase is explained by the presence of activation domains in the C-termini of naturally occurring and synthetic TALEs, which can induce expression of a gene when targeted to a promoter (Cermak et al., 2011; Streubel et al., 2017). Notably, the tNrp1-RBBP5 and -EZH2 fusions, which are larger than the tNrp1-JMJD3 fusion (Figure 6.14B) did not induce *Nrp1* expression. Given that the C-terminal end of JMJD3 confers a demethylase activity (Kruidenier et al., 2012), this result suggested that the increased expression observed was due to reduced levels of H3K27me3 at the *Nrp1* locus.

To test this, I performed ChIP-qPCR using an antibody that recognizes H3K27me3. Unfortunately, only one biological replicate of the H3K27me3 was successful and indicated that there was no decrease in H3K27me3 upon expression of tNrp1-JMJD3 (Figure 6.21). It is possible that the ChIP-qPCR was not sensitive enough or was somehow compromised. Indeed, it is suspicious that only two of the three IP samples from the ChIP replicates produced a detectable signal. It is also concerning that large differences were observed in H3K27me3 enrichment at the control genes between the tNrp1-JMJD3 and TALE-Only-A samples. I also performed a ChIP using an antibody that recognizes H3K4me3 in order to correlate any putative changes in methylation status at this histone tail with increases in expression associated with tNrp1-JMJD3 activity. Although all three biological replicates were successfully performed using this antibody, no changes in H3K4me3 status were observed. This is perhaps not a surprising result given that little or no effect on global gene expression levels are observed when global changes to H3K4me3 enrichment are induced (Clouaire et al., 2012; Hormanseder et al., 2017; Lenstra et al., 2011; Muramoto et al., 2010).

Certain reports have shown that removal of H3K27me3 by JMJD3 can facilitate the induction of transcriptional elongation (Chen et al., 2012; Estarás et al., 2013) or the loss of a transcriptionally repressed state over time (Coleman and Struhl, 2017). This would suggest that a loss in the H3K27me3 mark would occur before an increase in gene expression. In order to more completely understand the observed changes in expression, it would be desirable to repeat the H3K27me3 ChIP over a time-course to test whether *Nrp1* expression changes are correlated with changes in H3K27me3 enrichment. It is also possible that the increase observed in transcription of *Nrp1* following tNrp1-JMJD3 treatment was caused by a demethylase-independent function of JMJD3. It has been shown that JMJD3 also plays a role in activating gene transcription through chromatin remodelling (Miller et al., 2010). To test if the increase in *Nrp1* expression is independent of the demethylase capability of JMJD3, the effect of a tNrp1 fusion protein containing a catalytically inactive form of JMJD3 on *Nrp1* expression could be examined.

A central issue, which this study partly attempts to address, is whether histone status truly reflects or affects gene expression levels and *vice versa*. It has been reported that large increases in H3K4me3 at particular gene promoters only induce very small changes in gene expression (Cano-Rodriguez et al., 2016). These findings were reflected in this study by the lack of change in H3K4me3 status upon high levels of *Nrp1* induction by tNrp1-VP64, (Figure 6.19A, Figure 6.21A). It has also been shown that just a 15-fold increase in H3K4me3 enrichment at the human α-globin genes in hES cells vs. erythroid cells corresponds to a 30-40,000 times increase in α-globin gene expression (De Gobbi et al., 2011). However, as discussed in Chapter 1 (pp. 2), several lines of evidence have suggested that H3K4me3 may not be instructive of changes in gene expression, but may instead be more important in maintaining a memory of a transcriptional state (Clouaire et al., 2012; Hormanseder et al., 2017; Howe et al., 2017; Lenstra et al., 2011; Muramoto et al., 2010).

Nrp1 expression was unchanged after the transfection of cells with TALE fusions to RBBP5 or EZH2, which were hypothesized to stimulate H3K4me3 and H3K27me3 deposition, respectively (see section 6.1). However, it is important to note that previous reports suggest the targeting of multiple fusions along a DNA

sequence can improve the effects of these fusions on gene expression (Maeder et al., 2013b; Mali et al., 2013a; Perez-Pinera et al., 2013). It is possible that the tiling of the *Nrp1* promoter with multiple tNrp1-RBBP5 and –EZH2 fusions is required for a detectable effect on *Nrp1* gene expression. Indeed, a recent report showed that co-expression of three sgRNAs tiling a promoter and a dCas9-EZH2 fusion is sufficient to induce a 9-fold enrichment of H3K27me3 at that promoter (O'Geen et al., 2017). To more completely understand the performance of the tNrp1-RBBP5/EZH2 vectors, tiling experiments of the *Nrp1* promoter with these fusions should be performed. Both qPCR and ChIP-qPCRs should then be performed to assess the levels of *Nrp1* induction and H3K27me3 and H3K4me3 at the *Nrp1* locus. Furthermore, ChIP-qPCR should be performed to assay the presence or absence of other components associated with the activities of these enzymes, such as the MLL and PRC2 components.

Conclusions

I have successfully generated and validated an efficient cloning system that will facilitate future experiments where complex gene fusions must be created. I have also replicated experimental findings from the literature and added to that knowledge by showing that H3K4me3 levels are not elevated at the *Nrp1* locus after artificially inducing *Nrp1* expression. Furthermore, I have performed the first experiments aimed at interrogating the effects of modulating the presence of chemical groups on the tails of histones at the *Nrp1* locus. These experiments have led to some interesting results, such as the observation that the tNrp1-JMJD3 fusion protein promoted *Nrp1* expression at certain time-points, possibly by influencing H3K27me3 levels. However, as discussed in the above sections, more complete analyses are required to fully understand these results. Nevertheless, the advances described here have informed an improved experimental design. Progress will be more rapid in the second round of experiments as the cloning system has been generated and validated. Future experiments will include testing these TALE fusions at additional bivalent loci in ES cells, including the human α-globin genes, and monitoring changes in gene expression and histone modifications during *in vitro* differentiation of these cells.

Chapter 7 Discussion and Conclusions

Cellular differentiation depends on the activation and repression of specific gene sets at particular times during development. This is partially directed by epigenetic states, which are correlated with gene expression and silencing. Over the last decade, a body of evidence has shown that apparently antagonistic epigenetic marks can co-localise to the same locus (termed bivalent domains). Although the biological significance of this co-localisation for gene expression remains under investigation, the techniques used to assay these states often lack the sufficient spatial resolution to define these genes as bivalent. Using the human α-globin locus as a model for bivalency, I aimed to answer two questions during my thesis. The first was to understand the true nature of bivalency at this locus, as it was originally identified as a bivalent gene using techniques that cannot distinguish between bivalency and mixed cell populations. The second question aimed at understanding the function of bivalency during development, which has been proposed to support developmental phase transitions.

7.1 Characterising the nature of bivalency

A heterogeneous human α-globin expression pattern in humanised mouse erythroid cells was observed when the HS-40 enhancer was removed (ΔHS-40)(De Gobbi et al., 2017). Notably, bivalency did not resolve at the human α-globin genes in these ΔHS-40 mouse erythroid cells (Vernimmen et al., 2011). An epigenetically mixed population of cells may underlie the heterogeneous human α-globin expression pattern observed in the ΔHS-40 mouse erythroid cells. This would suggest that the PcG protein removal mediated by the HS-40 enhancer (Vernimmen et al., 2011) is important for the maintenance of an epigenetically homogenous population and for the resolution of bivalency at the human α-globin genes. To test this hypothesis, I aimed to generate an in-frame fusion of a *2A-mCherry* reporter cassette with the *HBA2* gene in the human α-globin locus in humanised mES cell lines. Erythroid cells could then be isolated from these mES cells by *in vitro* differentiation, or from primary cells in a mouse knock-in line. The mCherry

expressing and non-expressing populations could then be isolated by FACS and analysed separately by ChIP-qPCR.

In Chapter 3, I subcloned the humanised WTH and ΔHS-40 mES cell lines. I identified two subclones, WTH3 and ΔHS-40_20, with a suitable karyotype to perform gene targeting experiments. I also tested numerous variations of an *in vitro* differentiation protocol to generate erythroid cells and optimised a robust and efficient method for use with the WTH3 subclone. Upon completion of these critical steps, I designed gene-targeting strategies and generated a number of gene targeting tools in Chapter 4. I tailored a CRISPR/Cas9 gene targeting strategy and generated two HDR donor vectors while I also designed and validated a sgRNA targeting the *HBA2* 3`UTR. These HDR donor vectors were designed to introduce a short mutation to the *HBA2* 3`UTR, which was designed to prevent the sgRNA from recognising its target site after the insertion of the desired cassette. I confirmed that this 3`UTR mutation does not reduce reporter gene expression. Once all of the tools had been generated and validated, I began gene-targeting experiments using WTH3 mES cells in Chapter 5. The gene targeting experiments required multiple rounds of optimization before a number of correct knock-in clones with a single copy of a *2A-mCherry* insertion were obtained.

The WTH3 cell line containing a *2A-mCherry* insertion at *HBA2* generated in this study will be further engineered to remove the *PiggyBac* flanked selection cassette and the HS-40 enhancer. Transfection of the cells with a hyperactive transposase will mediate the removal of the selection cassette. SgRNAs targeting the 5` and 3` ends of the HS-40 enhancer have been designed and validated (data not shown). Transfection of the cells with a combination of these sgRNA constructs should mediate the deletion of the HS-40 enhancer region in the presence of Cas9. Erythroid cells could then be obtained from the WTH3-HBA2-2AmCherry and ΔHS-40-HBA2-2AmCherry mES cell lines by *in vitro* differentiation, or from primary cells from mouse knock-in lines. The mCherry expressing and non-expressing populations of cells could be isolated by FACS and analysed by ChIP qPCR.

Three experimental outcomes appear possible from these ChIP qPCR analyses. Firstly, it seems possible that the expression status of *HBA2* will be reflected by its epigenetic status in the ΔHS-40 mutant background. If so, the mCherry positive population of ΔHS-40 erythroid cells would be marked by H3K4me3, but not H3K27me3, at the *HBA2* gene. This data would suggest that the primary α-globin enhancer, HS-40, increases the probability of resolution of bivalency at its target gene, *HBA2*. Indeed, this enhancer-mediated resolution of bivalency at *HBA2* could be achieved *via* the recruitment of the histone demethylase JMJD3 by HS-40 (Vernimmen et al., 2011). Furthermore, if the mCherry negative population of ΔHS-40 erythroid cells were monovalently marked with H3K27me3 this would suggest that both histone marks are directly involved in controlling gene expression.

A second possible outcome would be that the mCherry positive population in the ΔHS-40 erythroid cells are monovalently marked with H3K4me3 while the mCherry negative population are bivalently marked with both H3K4me3 and H3K27me3. This observation would also suggest that the HS-40 enhancer increases the probability of H3K27me3 removal. Interestingly, this result would imply that the removal of the H3K27me3 mark is required for the activation of *HBA2* expression as no actively expressing cells contain this mark.

It is also possible that both the mCherry positive and negative population in the ΔHS-40 erythroid cells are bivalently marked with H3K4me3 and H3K27me3 at *HBA2*. This data would suggest that the presence of the HS-40 enhancer is necessary for the removal of the H3K27me3 mark at *HBA2* in any erythroid cell. However this would also show that the removal of the H3K27me3 mark is not required for the activation of *HBA2* gene expression.

Techniques that do not require the use of a fluorescent reporter knock-in cell line could also be used to characterise the nature of bivalency at the human α-globin genes in WTH and ΔHS-40 erythroid cells. As described in Chapter 1 (pp. 12), recent technical advancements have managed to map bivalency to genomic loci in a number of cell types (Shema et al., 2016; Weiner et al., 2016). Co-ChIP or single molecule imaging/sequencing could be used to determine if the H3K4me3 and

H3K27me3 marks co-localise to the same individual nucleosomes across the human α-globin genes in the ΔHS-40 erythroid cells. However, these techniques are limited as they would be unable to directly indicate whether each nucleosome (marked by H3K4me3 only, H3K27me3 only or both) originated from an α-globin expressing cell. Single-molecule imaging/sequencing would reveal the proportion of ΔHS-40 erythroid cells that carry each of the histone marks at the human α-globin gene. Our hypothesis suggests that the histone modification status at the human α-globin genes underlies human α-globin expression level in these cells; i.e. α-globin genes marked with H3K4me3 would be active and α-globin genes marked with both H3K4me3 and H3K27me3 or H3K27me3 alone would be inactive. Quantitative assessment of the ΔHS-40 erythroid cells has already shown that ~50% express human α-globin (De Gobbi et al., 2017). If the histone modification status reflects the binary α-globin expression status in individual cells, single molecule-imaging/sequencing data would reveal that ~50% of ΔHS-40 erythroid cells would be marked with H3K4me3 and the rest would be marked with both H3K4me3 and H3K27me3 or H3K27me3 alone.

Nano-flare technology, described in Chapter 1 (pp. 43), could also be adapted to facilitate the isolation of cells that express human α-globin without the use of gene editing. This technique relies on the uptake of a spherical nucleic acid probe bound to a gold nanoparticle (Seferos et al., 2007). These nanoparticles are taken into the cell *via* receptor-mediated endocytosis, the process by which important factors involved in appropriate differentiation are absorbed into developing erythroid cells (Iacopetta et al., 1983; Sawyer et al., 1987). Once the nano-flare is within the cell, the binding of the target RNA transcript releases a fluorophore, allowing the detection of cells that are transcribing a particular gene (Seferos et al., 2007). This technology could be optimised for use in mouse erythroblasts using a probe homologous to human α-globin in order to separate ΔHS-40 expressing and non-expressing cells. These two populations could then be analysed by ChIP-qPCR.

A human α-globin reporter system has not yet been published and would provide a very flexible tool for studying human α-globin transcription. The tools generated here and the candidate population of targeted cell lines lay a foundation for the generation of this reporter. The results of ChIP qPCR experiments on erythroid

cells generated from these reporter lines would provide a novel insight into how enhancers are involved in the resolution of epigenetic bivalency and further data on how histone marks are involved in controlling gene expression.

7.2 Characterising the function of bivalency

The significance of bivalency during development is currently unclear. As discussed in Chapter 1 (pp. 18), the global depletion of H3K4me2/3 at bivalent genes rarely leads to the disruption of gene regulation or differentiation capacity of these cells (Denissov et al., 2014; Hu et al., 2013; Rickels et al., 2016). Conversely, effects have been observed on gene regulation when the H3K27me3 mark is depleted from tissue-specific bivalent genes (Jadhav et al., 2016). It is important to understand if bivalency is a developmentally significant feature of the epigenetic landscape, and to understand the contribution of each of these marks to the function of bivalency. Modifying epigenetic statuses in a site-specific manner and analysing how this affects gene expression would help to answer these questions.

In Chapter 6 I developed a cloning strategy to generate TALE fusions to epigenetic editing enzymes. I used a TALE sequence targeting the bivalent mouse promoter *Nrp1* and fused this to RBBP5, EZH2 and the catalytic domain of JMJD3. Interestingly, the tNrp1-JMJD3 fusion, containing the catalytic domain of the H3K27me3 histone demethylase JMJD3, caused a mild increase in *Nrp1* gene expression. However, changes in the H3K4me3/H3K27me3 mark enrichment were undetectable by ChIP-qPCR. Many previous reports have suggested that the targeting of multiple fusions along a DNA sequence can improve the effects of these fusions on gene expression and histone modifications (Maeder et al., 2013b; Mali et al., 2013a; Perez-Pinera et al., 2013). It therefore seems likely that the use of additional tNrp1-JMJD3 fusions tiling the *Nrp1* promoter may cause a greater effect on *Nrp1* gene expression and histone modification status. It is also possible that tiling of the *Nrp1* promoter with the tNrp1-RBBP5 and –EZH2 fusions is required for a detectable effect on *Nrp1* gene expression. Indeed, these preliminary experiments provided a solid basis for future work and a number of possible improvements to the

experimental strategy are discussed in Chapter 6. The use of epigenetic editing tools can help to elucidate the function of individual histone modifications during transcriptional regulation. Furthermore, the manipulation of H3K4me3 and H3K27me3 at human α-globin (and other genes) in WTH and ΔHS-40 mES cells, followed by the analysis of dynamic changes in gene expression during *in vitro* differentiation, will reveal information related to the function of bivalency during development. Several other questions relating to the function of specific epigenetic states can also be addressed using these epigenetic editing tools. For example, it is unknown how the addition of the H3K27me3 mark would affect gene transcription during differentiation at genes that are monovalently marked with H3K4me3 but not expressed (Vastenhouw et al., 2010). It would also be interesting to assess the effects of manipulating the H3K4me3 mark at bivalent genes and monovalent H3K27me3 marked genes. The epigenetic editing tools generated in this thesis provide a solid basis with which to investigate a variety of questions in a site-specific manner.

This thesis has generated a number of tools for genetic editing and epigenetic editing at bivalent genes. These tools and experiments provide an important foundation for future experiments investigating both the nature of bivalency at the human α-globin locus and the function of bivalency during development.

Chapter 8 Bibliography

Agger, K., Cloos, P. A. C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A. E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature *449*, 731–734.

Agger, K., Christensen, J., Cloos, P. A. C., and Helin, K. (2008). The emerging functions of histone demethylases. Curr. Opin. Genet. Dev. *18*, 159–168.

Ahmadiyeh, N., Pomerantz, M. M., Grisanzio, C., Herman, P., Jia, L., and Almendro, V. (2010). 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proc. Natl. Acad. Sci. U. S. A. *107*, 9742–9746.

Akhtar, W., de Jong, J., Pindyurin, A. V, Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L. F. A., van Lohuizen, M., and van Steensel, B. (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell *154*, 914–927.

Akhtar-Zaidi, B., Cowper-Sallari, R., Corradin, O., Saiakhova, A., Bartels, C. F., Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M. F., et al. (2012). Epigenomic enhancer profiling defines a signature of colon cancer. Science *336*, 736–739.

Al-Hasani, K., Vadolas, J., Voullaire, L., Williamson, R., and Ioannou, P. A. (2004). Complementation of alpha-thalassaemia in alpha-globin knockout mice with a 191 kb transgene containing the human alpha-globin locus. Transgenic Res. *13*, 235–243.

Albitar, M., Cash, F.E., Peschle, C., and Liebhaber, S. A. (1992). Developmental switch in the relative expression of the alpha 1- and alpha 2-globin genes in humans and in transgenic mice. Blood *79*, 2471–2474.

Amabile, A., Migliara, A., Capasso, P., Biffi, M., Cittaro, D., Naldini, L., and Lombardo, A. (2016). Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. Cell *167*, 219–232.e14.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. Dev. Cell *16*, 47–57.

Anastassiadis, K., Fu, J., Patsch, C., Hu, S., Weidlich, S., Duerschke, K., Buchholz, F., Edenhofer, F., and Stewart, A. F. (2009). Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice. Dis. Model. Mech. *2*, 508–515.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014a). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Andersson, R., Refsing Andersen, P., Valen, E., Core, L. J., Bornholdt, J., Boyd, M., Heick Jensen, T., and Sandelin, A. (2014b). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. Nat. Commun. *5*, 5336.

Anguita, E., Hughes, J., Heyworth, C., Blobel, G. A., Wood, W.G., and Higgs, D. R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. EMBO J. *23*, 2841–2852.

Arcipowski, K. M., Martinez, C. A., and Ntziachristos, P. (2016). Histone demethylases in physiology and cancer: A tale of two enzymes, JMJD3 and UTX. Curr. Opin. Genet. Dev. *36*, 59–67.

Arner, E., Daub, C., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rönnerblad, M., Hrydziuszko, O., Vitezic, M., et al. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science. *347*, 1010–1015.

Arnold, P., Scho, A., Pachkov, M., Balwierz, P. J., Jørgensen, H., Stadler, M. B., Nimwegen, E. Van, and Schu, D. (2013). Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. Genome Res. 60–73.

Aspalter, I. M., Gordon, E., Dubrac, A., Ragab, A., Narloch, J., Vizán, P., Geudens, I., Collins, R. T., Franco, C. A., Abrahams, C. L., et al. (2015). Alk1 and Alk5 inhibition by Nrp1 controls vascular sprouting downstream of Notch. Nat. Commun. *6*, 7264.

Avdic, V., Zhang, P., Lanouette, S., Groulx, A., Tremblay, V., Brunzelle, J., and Couture, J. F. (2011). Structural and biochemical insights into MLL1 core complex assembly. Structure *19*, 101–108.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. *8*, 532–538.

Baker, O., Tsurkan, S., Fu, J., Klink, B., Rump, A., Obst, M., Kranz, A., Schröck, E., Anastassiadis, K., and Stewart, A. F. (2017). The contribution of homology arms to nuclease-assisted genome engineering. Nucleic Acids Res. 1–11.

Bakstad, D., Adamson, A., Spiller, D. G., and White, M. R. H. (2012). Quantitative measurement of single cell dynamics. Curr. Opin. Biotechnol. *23*, 103–109.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Bannister, A. J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. J. Biol. Chem. *280*, 17732–17736.

Bartman, C. R., Hsu, S. C., Hsiung, C. C. S., Raj, A., and Blobel, G. A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol. Cell *62*, 237–247.

Bassett, A. R., Tibbit, C., Ponting, C. P., and Liu, J. (2013). Resource Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR / Cas9 System. Cell Reports *4*, 220–228.

Becker, J. S., Nicetto, D., and Zaret, K. S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet. *32*, 29–41.

Bell, C. C., Magor, G. W., Gillinder, K. R., and Perkins, A. C. (2014). A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing. BMC Genomics *15*, 1002.

Benayoun, B. A, Pollina, E. A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E. D., Devarajan, K., Daugherty, A. C., Kundaje, A. B., Mancini, E., et al. (2014). H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell *158*, 673–688.

Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. Genes Dev. *23*, 781–783.

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., Gingeras, T. R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169–181.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048.

Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Bernt, K. M., Zhu, N., Sinha, A. U., Vempati, S., Faber, J., Krivtsov, A. V., Feng, Z., Punt, N., Daigle, A., Bullinger, L., et al. (2011). MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. Cancer Cell *20*, 66–78.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H., and Long, R.M. (1998). Localization of ASH1 mRNA Particles in Living Yeast. Mol. Cell *2*, 437–445.

Biagioli, M., Pinto, M., Cesselli, D., Zaninello, M., Lazarevic, D., Roncaglia, P.,

Simone, R., Vlachouli, C., Plessy, C., Bertin, N., et al. (2009). Unexpected expression of alpha- and beta-globin in mesencephalic dopaminergic neurons and glial cells. Proc. Natl. Acad. Sci. U. S. A. *106*, 15454–15459.

Black, J. B., Adler, A. F., Wang, H. G., D'Ippolito, A. M., Hutchinson, H. A., Reddy, T. E., Pitt, G. S., Leong, K. W., and Gersbach, C. A. (2016). Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. Cell Stem Cell *19*, 406–414.

Blackledge, N. P., Thomson, J. P., and Skene, P. J. (2013). CpG island chromatin is shaped by recruitment of ZF-CxxC proteins. Cold Spring Harb. Perspect. Biol. *5*, a018648.

Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L. P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. Cell *157*, 1445–1459.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509–1512.

Boettcher, M., and McManus, M. T. (2015). Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. Mol. Cell *58*, 575–585.

Boheler, K. R., Czyz, J., Tweedie, D., Yang, H. T., Anisimov, S. V., and Wobus, A. M. (2002). Differentiation of pluripotent embryonic stem cells into cardiomyocytes. Circ. Res. *91*, 189–201.

Bondurand, N., Fouquet, V., Baral, V., Lecerf, L., Loundon, N., Goossens, M., Duriez, B., Labrune, P., and Pingault, V. (2012). Alu-mediated deletion of SOX10 regulatory elements in Waardenburg syndrome type 4. Eur. J. Hum. Genet. *20*, 990–994.

Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H., and Helin, K. (2006). Roles in cell fate transitions Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. 1123–1136.

Branda, C. S., and Dymecki, S. M. (2004). Talking about a Revolution: The Impact of Site-Specific Recombinases on Genetic Analyses in Mice. Dev. Cell *6*, 7–28.

Brazel, A. J., and Vernimmen, D. (2016). The complexity of epigenetic diseases. J. Pathol. *238*, 333–344.

Brind'Amour, J., Liu, S., Hudson, M., Chen, C., Karimi, M. M., and Lorincz, M. C. (2015). An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. Nat. Commun. *6*, 6033.

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. RNA *19*, 429–

442.

Brookes, E., Santiago, D., Hebenstreit, D., Morris, K. J., Carroll, T., Xie, S. Q., Stock, J. K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs. Cell Stem Cell *10*, 157–170.

Burgold, T., Spreafico, F., De Santa, F., Totaro, M. G., Prosperini, E., Natoli, G., and Testa, G. (2008). The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment. PLoS One *3*, e3034.

Cai, M., Kim, S., Wang, K., Farnham, P. J., Coetzee, G. A., and Lu, W. (2016). 4C-seq revealed long-range interactions of a functional enhancer at the 8q24 prostate cancer risk locus. Sci. Rep. *6*, 22462.

Cano-Rodriguez, D., Gjaltema, R. A. F., Jilderda, L. J., Jellema, P., Dokter-Fokkens, J., Ruiters, M. H. J., and Rots, M. G. (2016). Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. Nat. Commun. *7*, 12284.

Cantone, I., and Fisher, A. G. (2013). Epigenetic programming and reprogramming during development. Nat. Sructural Mol. Biol. *20*, 282–289.

Cao, P., Deng, Z., Wan, M., Huang, W., Cramer, S. D., Xu, J., and Lei, M. (2010). MicroRNA-101 negatively regulates Ezh2 and its expression is modulated by androgen receptor and HIF-1alpha/HIF-1beta. Mol. Cancer *9*.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. Science. *298*, 1039–1043.

Cao, R., Tsukada, Y., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing. Mol. Cell *20*, 1–4.

Capecchi, M. R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat. Rev. Genet. *6*, 507–512.

Carotta, S., Pilat, S., Mairhofer, A., Schmidt, U., Dolznig, H., Steinlein, P., and Beug, H. (2004). Directed differentiation and mass cultivation of pure erythroid progenitors from mouse embryonic stem cells. Blood *104*, 1873–1881.

Carroll, D. O., Erhardt, S., Pagani, M., and Barton, S. C. (2001). The Polycomb - Group Gene Ezh2 Is Required for Early Mouse Development. Mol. Cell. Biol. *21*, 4330–4336.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. *10*, 295–304.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. a, Somia, N. V, Bogdanove, A. J., and Voytas, D. F. (2011). Efficient design and

assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. *39*, e82.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chapman, J. R., Taylor, M. R. G., and Boulton, S. J. (2012). Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. Mol. Cell *47*, 497–510.

Charlet, J., Duymich, C. E., Lay, F. D., Mundbjerg, K., Dalsgaard Sørensen, K., Liang, G., and Jones, P. A. (2016). Bivalent Regions of Cytosine Methylation and H3K27 Acetylation Suggest an Active Role for DNA Methylation at Enhancers. Mol. Cell *62*, 422–431.

Chen, S., Ma, J., Wu, F., Xiong, L. J. J., Ma, H., Xu, W., Lv, R., Li, X., Villen, J., Gygi, S. P., et al. (2012). The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation. Genes Dev. *26*, 1364–1375.

Choi, P. S., and Meyerson, M. (2014). Targeted genomic rearrangements using CRISPR/Cas technology. Nat. Commun. *5*, 3728.

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., and Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics *186*, 756–761.

Clark, B., Shooter, C., Smith, F., Brawand, D., Steedman, L., Oakley, M., Rushton, P., Rooks, H., Wang, X., Drousiotou, A., et al. (2016). Beta thalassaemia intermedia due to co-inheritance of three unique alpha globin cluster duplications characterised by next generation sequencing analysis. Br. J. Haematol.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J. H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes Dev. *26*, 1714–1728.

Coelho, A., Picanço, I., Seuanes, F., Seixas, M. T., and Faustino, P. (2010). Novel large deletions in the human alpha-globin gene cluster: Clarifying the HS-40 long-range regulatory role in the native chromosome environment. Blood Cells, Mol. Dis. *45*, 147–153.

Coleman, R. T., and Struhl, G. (2017). Causal role for inheritance of H3K27me3 in maintaining the OFF state of a *Drosophila* HOX gene. Science (80-.). *356*, eaai8236.

Le Cong, Ran, F. A., Cox, D., Shuailiang, L., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. *339*, 819–823.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N. P.,

DeMarco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. Cell Rep. *7*, 1456–1470.

Court, F., and Arnaud, P. (2017). An annotated list of bivalent chromatin regions in human ES cells: a new tool for cancer epigenetic research. Oncotarget *8*, 4110–4124.

Cradick, T. J., Fine, E. J., Antico, C. J., and Bao, G. (2013). CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res. *41*, 9584–9592.

Creyghton, M. P., Cheng, A. W., Welstead, G.G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K. J., Gooding, A. R., Lee, J. T., and Cech, T. R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. Mol. Cell *57*, 552–558.

Deaton, A., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A., and Blobel, G. A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233–1244.

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H. G., and Stewart, A. F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Development *141*, 526–537.

Dever, D. P., Bak, R. O., Reinisch, A., Camarena, J., Washington, G., Nicolas, C. E., Pavel-Dinu, M., Saxena, N., Wilkens, A. B., Mantri, S., et al. (2016). CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. Nature *539*, 384–389.

Dhar, S. S., Lee, S. H., Chen, K., Zhu, G., Oh, W. K., Allton, K., Gafni, O., Kim, Y. Z., Tomoiga, A. S., Barton, M. C., et al. (2016). An essential role for UTX in resolution and activation of bivalent promoters. Nucleic Acids Res. *44*, 3659–3674.

van Dijk, T. B., Gillemans, N., Stein, C., Fanis, P., Demmers, J., van de Corput, M., Essers, J., Grosveld, F., Bauer, U. M., and Philipsen, S. (2010). Friend of Prmt1, a novel chromatin target of protein arginine methyltransferases. Mol. Cell. Biol. *30*, 260–272.

Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene

in mouse embryonic stem cells. Nature *328*, 429–432.

Dorschner, M. O., Hawrylycz, M., Humbert, R., Wallace, J. C., Shafer, A., Kawamoto, J., Mack, J., Hall, R., Goldy, J., Sabo, P. J., et al. (2004). High-throughput localization of functional elements by quantitative chromatin profiling. Nat. Methods *1*, 219–225.

Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat. Struct. Mol. Biol. *13*, 713–719.

Doyle, E. L., Booher, N. J., Standage, D. S., Voytas, D. F., Brendel, V. P., Vandyk, J. K., and Bogdanove, A. J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: Tools for TAL effector design and target prediction. Nucleic Acids Res. *40*, 117–122.

Dubridge, R. B., Tang, P., Hsia, H. A. N. C., Leong, P., Miller, J. H., and Calos, M. P. (1987). Analysis of Mutation in Human Cells by Using an Epstein-Barr Virus Shuttle System. Mol. Cell. Biol. *7*, 379–387.

Dzierzak, E., and Philipsen, S. (2013). Erythropoiesis : Development and Differentiation. Cold Spring Harb Perspect Med. *3*, a011601.

Eberl, H. C., Spruijt, C. G., Kelstrup, C. D., Vermeulen, M., and Mann, M. (2013). A Map of General and Specialized Chromatin Readers in Mouse Tissues Generated by Label-free Interaction Proteomics. Mol. Cell *49*, 368–378.

Emison, E. S., Mccallion, A. S., Kashuk, C. S., Bush, R. T., Grice, E., Lin, S., Portnoy, M. E., Cutler, D. J., Green, E. D., and Chakravarti, A. (2005). A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. Nature *434*, 857–863.

Engelke, D. R., Krikos, A., Bruck, M. E., and Ginsburg, D. (1990). Purification of Thermus aquaticus DNA polymerase expressed in Escherichia coli. Anal. Biochem. *191*, 396–400.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*, e3647.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. PLoS One *4*.

Estarás, C., Fueyo, R., Akizu, N., Beltrán, S., and Martínez-Balbás, M. A. (2013). RNA polymerase II progression through H3K27me3-enriched gene bodies requires JMJD3 histone demethylase. Mol. Biol. Cell *24*, 351–360.

Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotent cells from mouse embryos. Nature *292*, 154–156.

Fantin, A., Vieira, J. M., Plein, A., Denti, L., Fruttiger, M., Pollard, J. W., and Ruhrberg, C. (2013). NRP1 acts cell autonomously in endothelium to promote tip cell function during sprouting angiogenesis. Blood *121*, 2352–2362.

Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., Lee, S., Sims, D., Cerase, A., Sheahan, T. W., et al. (2012). KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. Elife *2012*, 1–26.

Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S., and Levine, M. S. (2015). Suboptimization of developmental enhancers. Science *350*, 325–328.

Farley, E. K., Olson, K. M., Zhang, W., Rokhsar, D. S., and Levine, M. S. (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. Proc. Natl. Acad. Sci. U. S. A. *113*, 6508–6513.

Faust, C., Lawson, K. A., Schork, N. J., Thiel, B., and Magnuson, T. (1998). The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. Development *125*, 4495–4506.

Femino, A. M., Fay, F. S., Fogarty, K., and Singer, R. H. (1998). Visualization of single RNA transcripts in situ. Science *280*, 585–590.

Ferreira, R., Ohneda, K., Yamamoto, M., and Philipsen, S. (2005). GATA1 Function, a Paradigm for Transcription Factors in Hematopoiesis. Mollecular Cell. Biol. *25*, 1215–1227.

Fiering, S., Whitelaw, E., and Martin, D.I . (2000). To be or not to be active: the stochastic nature of enhancer action. Bioessays *22*, 381–387.

Figueroa, M. E., Lugthart, S., Li, Y., Erpelinck-Verschueren, C., Deng, X., Christos, P.J., Schifano, E., Booth, J., van Putten, W., Skrabanek, L., et al. (2010). DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell *17*, 13–27.

Forget, B. G., and Hardison, R. C. (2009). The Normal Structure and Regulation of Human Globin Gene Clusters. Disord. Hemoglobins Genet. Pathophysiol. Clin. Manag. 46–61.

Francis, N. J., Kingston, R.E., and Woodcock, C. L. (2004). Chromatin Compaction by a Polycomb Group Protein Complex. Science. *306*, 1574–1577.

Frankel, N., Erezyilmaz, D. F., McGregor, A. P., Wang, S., Payre, F., and Stern, D. L. (2011). Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. Nature *474*, 598–603.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358–368.

Fukuda, H., Takahashi, J., Watanabe, K., Hayashi, H., Morizane, A., Koyanagi, M.,

Sasai, Y., and Hashimoto, N. (2006). Fluorescence-Activated Cell Sorting–Based Purification of Embryonic Stem Cell–Derived Neural Precursors Averts Tumor Formation after Transplantation. Stem Cells *24*, 763–771.

Fullerton, S. M., Bernardo Carvalho, A., and Clark, A. G. (2001). Local rates of recombination are positively correlated with GC content in the human genome. Mol Biol Evol *18*, 1139–1142.

Gao, X., Yang, J., Tsang, J. C. H., Ooi, J., Wu, D., and Liu, P. (2013). Reprogramming to pluripotency using designer TALE transcription factors targeting enhancers. Stem Cell Reports *1*, 183–197.

Gaztelumendi, N., and Nogués, C. (2014). Chromosome instability in mouse embryonic stem cells. Sci. Rep. *4*, 5324.

Ghamari, A., van de Corput, M. P. C., Thongjuea, S., van Cappellen, W. A., van IJcken, W., van Haren, J., Soler, E., Eick, D., Lenhard, B., and Grosveld, F.G. (2013). In vivo live imaging of RNA polymerase II transcription factories in primary cells. Genes Dev. 767–777.

Giarratana, M., Rouard, H., Dumont, A., Kiger, L., Safeukui, I., Le Pennec, P.Y., Francois, S., Trugnan, G., Peyrard, T., Marie, T., et al. (2011). Proof of principle for transfusion of *in vitro* generated red blood cells. Blood *118*, 5071–5079.

Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Hutchison III, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

De Gobbi, M., Garrick, D., Lynch, M., Vernimmen, D., Hughes, J. R., Goardon, N., Luc, S., Lower, K. M., Sloane-Stanley, J. A., Pina, C., et al. (2011). Generation of bivalent chromatin domains during cell fate decisions. Epigenetics Chromatin *4*, 9.

De Gobbi, M., Brazel, A. J., Sharpe, J. A., Sloane-Stanley, J. A., Smith, A. J., Wood, W. G., and Vernimmen, D. (2017). Enhancer deletion generates cellular phenotypic diversity due to bimodal gene expression. Blood Cells, Mol. Dis. *64*, 10–12.

Gómez-Rodríguez, J., Washington, V., Cheng, J., Dutra, A., Pak, E., Liu, P., Mcvicar, D.W., and Schwartzberg, P. L. (2008). Advantages of q-PCR as a method of screening for gene targeting in mammalian cells using conventional and whole BAC-based constructs. Nucleic Acids Res. *36*, 1–9.

Gonzalez, F. (2016). CRISPR/Cas9 genome editing in human pluripotent stem cells: Harnessing human genetics in a dish. Dev. Dyn. *245*, 788–806.

Gossler, A., Doetschman, T., Korn, R., Serflingt, E., and Kemler, R. (1986).

Transgenesis by means of blastocyst-derived embryonic stem cell lines. Dev. Biol. *83*, 9065–9069.

Grampp, S., Platt, J. L., Lauer, V., Salama, R., Kranz, F., Neumann, V. K., Wach, S., Stöhr, C., Hartmann, A., Eckardt, K. U., et al. (2016). Genetic variation at the 8q24.21 renal cancer susceptibility locus affects HIF binding to a MYC enhancer. Nat. Commun. *7*, 13183.

Grandy, R. A., Whitfield, T. W., Wu, H., Fitzgerald, M. P., VanOudenhove, J. J., Zaidi, S. K., Montecino, M. A., Lian, J. B., van Wijnen, A. J., Stein, J. L., et al. (2015). Genome-Wide Studies Reveal that H3K4me3 Modification in Bivalent Genes Is Dynamically Regulated during the Pluripotent Cell Cycle and Stabilized upon Differentiation. Mol. Cell. Biol. *36*, 615–627.

Grimholt, R. M., Urdal, P., Klingenberg, O., and Piehler, A. P. (2014). Rapid and reliable detection of α-globin copy number variations by quantitative real-time PCR. BMC Hematol. *14*, 4.

Groner, A. C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Denervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS Genet. *6*.

Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell *51*, 975–985.

Guo, G., Pinello, L., Han, X., Lai, S., Shen, L., Lin, T. W., Zou, K., Yuan, G. C., and Orkin, S. H. (2016). Serum-Based Culture Conditions Provoke Gene Expression Variability in Mouse Embryonic Stem Cells as Revealed by Single-Cell Analysis. Cell Rep. *14*, 956–965.

Hall, D. B., and Struhl, K. (2002). The VP16 Activation Domain Interacts with Multiple Transcriptional Components as Determined by Protein-Protein Cross-linking *in Vivo*. J. Biol. Chem. *277*, 46043–46050.

Harteveld, C. L., Refaldi, C., Cassinerio, E., Cappellini, M. D., and Giordano, P. C. (2008). Segmental duplications involving the alpha-globin gene cluster are causing beta-thalassemia intermedia phenotypes in beta-thalassemia heterozygous patients. Blood Cells, Mol. Dis. *40*, 312–316.

Hatton, C. S., Wilkie, A. O., Drysdale, H., Wood, M. A., Vickers, J., Sharpe, J., Ayyub, H., Pretorius, I. M., Buckle, V. J., and Higgs, D. R. (1990). Alpha-Thalassemia Caused by a Large (62kb) Deletion Upstream of the Human alpha Globin Gene Cluster. Blood *76*, 221–227.

Hatzis, P., and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol. Cell *10*, 1467–1477.

Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L. L. P.,

Kassouf, M. T., Oudelaar, A. M., Sharpe, J. A., Suciu, M. C., et al. (2016). Genetic dissection of the α-globin super-enhancer *in vivo*. Nat. Genet. *48*, 895–903.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Herz, H. M., Garruss, A., and Shilatifard, A. (2013). SET for life: biochemical activities and biological functions of SET domain-containing proteins. Trends Biochem. Sci. *38*, 621–639.

Higgs, D. R. (2013). The molecular basis of α-thalassemia. Cold Spring Harb. Perspect. Med. *3*, a011718.

Higgs, D. R., and Wood, W. G. (2008). Long-range regulation of α globin gene expression during erythropoiesis. Curr. Opin. Hematol. *15*, 176–183.

Higgs, D. R., Vernimmen, D., and Wood, B. (2008). Long-Range Regulation of alpha -globin Gene Expression. Adv. Genet. *61*, 144–173.

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., and Gersbach, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Hitchins, M. P. (2015). Constitutional epimutation as a mechanism for cancer causality and heritability? Nat. Rev. Cancer *15*, 625–634.

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A., and Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hong, S., Cho, Y. W., Yu, L. R., Yu, H., Veenstra, T. D., and Ge, K. (2007). Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. Sci. U. S. A. *104*, 18439–18444.

Hormanseder, E., Simeone, A., Allen, G. E., Bradshaw, C. R., Figlmuller, M., Gurdon, J., and Jullien, J. (2017). H3K4 Methylation-Dependent Memory of Somatic Cell Identity Inhibits Reprogramming and Development of Nuclear Transfer Embryos. Cell Stem Cell *21*, 135–143.

Hosogane, M., Funayama, R., Nishida, Y., Nagashima, T., and Nakayama, K. (2013). Ras-induced changes in H3K27me3 occur after those in transcriptional activity. PLoS Genet. *9*, e1003698.

Howe, F. S., Fischl, H., Murray, S. C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? BioEssays *39*, 1–12.

Hsu, P. D., Scott, D. A., Weinstein, J. a, Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Hu, D., Garruss, A. S., Gao, X., Morgan, M. A., Cook, M., Smith, E. R., and Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1093–1097.

Huang, Y., Fang, J., Bedford, M. T., Zhang, Y., and Xu, R. M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science *312*, 748–751.

Hughes, J. R., Cheng, J., Ventress, N., Prabhakar, S., Clark, K., Anguita, E., De Gobbi, M., Jong, P., Rubin, E., and Higgs, D. R. (2005). Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. PNAS *102*.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J. R. J., and Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. *31*, 227–229.

Iacopetta, B. J., Morgan, E. H., and Yeoh, G. C. T. (1983). Receptor-mediated Endocytosis of Transferrin by Developing Erythroid Cells from the Rat Fetal Liver. J. Histochem. Cytochem. *31*, 336–344.

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. *41*, 178–186.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol. Cell. Biol. *24*, 8862–8871.

Jadhav, U., Nalapareddy, K., Saxena, M., O'Neill, N. K., Pinello, L., Yuan, G. C., Orkin, S. H., and Shivdasani, R. A. (2016). Acquired tissue-specific promoter bivalency is a basis for PRC2 necessity in adult cells. Cell *165*, 1389–1400.

Ji, X., Kong, J., Carstens, R. P., and Liebhaber, S. A. (2007). The 3` untranslated region complex involved in stabilization of human alpha-globin mRNA assembles in the nucleus and serves an independent role as a splice enhancer. Mol. Cell. Biol. *27*, 3290–3302.

Ji, X., Kong, J., and Liebhaber, S. A. (2011). An RNA-protein complex links enhanced nuclear 3` processing with cytoplasmic mRNA stabilization. EMBO J. *30*, 2622–2633.

Jia, J., Zheng, X., Hu, G., Cui, K., Zhang, J., Zhang, A., Jiang, H., Lu, B., Iii, J. Y., Liu, C., et al. (2012). Regulation of Pluripotency and Self- Renewal of ESCs through Epigenetic-Threshold Modulation and mRNA Pruning. Cell *151*, 576–589.

Jia, L., Landan, G., Pomerantz, M., Jaschek, R., Herman, P., Reich, D., Yan, C., Khalid, O., Kantoff, P., Oh, W., et al. (2009). Functional enhancers at the gene-poor 8q24 cancer-linked locus. PLoS Genet. *5*, e1000597.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816–821.

Joseph Hauptman (2009). Heat Inactivation — Are You Wasting Your Time ? Art To Sci. *15*, 1–4.

Joung, J. K., and Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. Nat. Rev. Mol. Cell Biol. *14*, 49–55.

Juven-Gershon, T., and Kadonaga, J. T. (2010). Regulation of gene expression *via* the core promoter and the basal transcriptional machinery. Dev. Biol. *339*, 225–229.

Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., Umezawa, A. and Reinberg, D. (2014) Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. Mol. cell, *53*(2), 290-300.

Kassis, J. A, and Brown, J. L. (2013). Polycomb group response elements in Drosophila and vertebrates. Adv. Genet. *81*, 83–118.

Kawasaki, T., Kitsukawa, T., Bekku, Y., Matsuda, Y., Sanbo, M., Yagi, T., and Fujisawa, H. (1999). A requirement for neuropilin-1 in embryonic vessel formation. Development *126*, 4895–4902.

Kearns, N. A., Pham, H., Tabak, B., Genga, R. M., Silverstein, N. J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat. Methods *12*.

Keller, G., Kennedy, M., Papayannopoulou, T., and Wiles, M. V. (1993). Hematopoietic commitment during embryonic stem cell differentiation in culture. Mol. Cell Biol. *13*, 473–486.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Hausler, D. (2002). The Human Genome Browser at UCSC. Genome Res. *12*, 996–1006.

Kidder, B. L., Hu, G., Yu, Z. X., Liu, C., and Zhao, K. (2013). Extended self-renewal and accelerated reprogramming in the absence of Kdm5b. Mol. Cell Biol. *33*, 4793–4810.

Kidder, B. L., Hu, G., and Zhao, K. (2014). KDM5B focuses H3K4 methylation near promoters and enhancers during embryonic stem cell self-renewal and

differentiation. Genome Biol. *15*, R32.

Kieffer-Kwon, K. R., Tang, Z., Mathe, E., Qian, J., Sung, M. H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell *155*, 1507–1520.

Kihm, A.J., Kong, Y., Hong, W., Russell, J. E., Rouda, S., Adachi, K., Simon, M. C., Blobel, G. A., and Weiss, M. (2002). An abundant erythroid protein that stabilizes free alpha-haemoglobin. Nature *417*, 758–763.

Kim, T. K., and Eberwine, J. H. (2010). Mammalian cell transfection: The present and the future. Anal. Bioanal. Chem. *397*, 3173–3178.

Kim, Y. M., Lee, J., Xia, L., Mulvihill, J. J., and Li, S. (2013). Trisomy 8: A common finding in mouse embryonic stem (ES) cell lines. Mol. Cytogenics *6*, 3–7.

Kimura, Y., Hisano, Y., Kawahara, A., and Higashijima, S. (2014). Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. Sci. Rep. *4*, 6545.

Kina, T., Ikuta, K., Takayama, E., Wada, K., Majumdar, A. Sen, Weissman, I. L., and Katsura, Y. (2000). The monoclonal antibody TER-119 recognizes a molecule associated with glycophorin A and specifically marks the late stages of murine erythroid lineage. Br. J. Haematol. *109*, 280–287.

Kinkley, S., Helmuth, J., Polansky, J. K., Dunkel, I., Gasparoni, G., Fröhler, S., Chen, W., Walter, J., Hamann, A., and Chung, H. R. (2016). reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4(+) memory T cells. Nat. Commun. *7*, 12514.

Kioussis, D., Vannin, E., DeLange, T., Flavell, R. A., and Grosveld, F. G. (1983). Beta-Globin gene inactivation by DNA translocation in gamma Beta-thalassaemia. Nature *306*, 662–666.

Klann, T. S., Black, J. B., Chellappan, M., Safi, A., Song, L., Hilton, I. B., Crawford, G. E., Reddy, T. E., and Gersbach, C. A. (2017). CRISPR – Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. Nat. Biotechnol. *35*, 561–568.

Klose, R. J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. Nature *442*, 312–316.

Klose, R. J., Cooper, S., Farcas, A. M., Blackledge, N. P., and Brockdorff, N. (2013). Chromatin sampling-an emerging perspective on targeting polycomb repressor proteins. PLoS Genet. *9*, 1–8.

Konermann, S., Brigham, M. D., Trevino, A. E., Hsu, P. D., Heidenreich, M., Cong, L., Platt, R. J., Scott, D. A., Church, G. M., and Zhang, F. (2013). Optical control of

mammalian endogenous transcription and epigenetic states. Nature *500*, 472–476.

Konermann, S., Brigham, M. D., Trevino, A .E., Joung, J., Abudayyeh, O. O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H., et al. (2014). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Kong, L., Tan, L., Lv, R., Shi, Z., Xiong, L., Wu, F., Rabidou, K., Smith, M., He, C., Zhang, L., et al. (2016). A primary role of TET proteins in establishment and maintenance of De Novo bivalency at CpG islands. Nucleic Acids Res. *44*, 8682–8692.

Kremers, G., Gilbert, S. G., Cranfill, P. J., Davidson, M. W., and Piston, D. W. (2011). Fluorescent proteins at a glance. J. Cell Sci. *124*, 157–160.

Krivtsov, A. V., and Armstrong, S. A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. Nat. Rev. Cancer *7*, 823–833.

Krivtsov, A. V., Feng, Z., Lemieux, M. E., Faber, J., Vempati, S., Sinha, A. U., Xia, X., Jesneck, J., Bracken, A. P., Silverman, L. B., et al. (2008). H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell *14*, 355–368.

Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D. P., Beattie, B. K., Emili, A., Boone, C., et al. (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol. Cell. Biol. *23*, 4207–4218.

Kruidenier, L., Chung, C., Cheng, Z., Liddle, J., Che, K., Joberty, G., Bantscheff, M., Bountra, C., Bridges, A., Diallo, H., et al. (2012). A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. Nature *488*, 404–408.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. *4*, e1000242.

Kurosawa, H. (2007). Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells. J. Biosci. Bioeng. *103*, 389–398.

Lakshmipathy, U., Pelacho, B., Sudo, K., Linehan, J., Coucouvanis, E., Kaufman, D.S., and Verfaillie, C. (2004). Efficient Transfection of Embryonic and Adult Stem Cells. Stem Cells *22*, 531–543.

Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A., and Singer, R. H. (2011). Real-time observation of transcription tnitiation and elongation on an endogenous yeast gene. Science. *332*, 475–478.

Lauberth, S. M., Nakayama, T., Wu, X., Ferris, A. L., Tang, Z., Hughes, S. H., and Roeder, R. G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation

complex assembly and selective gene activation. Cell *152*, 1021–1036.

Lecerf, L., Kavo, A., Ruiz-Ferrer, M., Baral, V., Watanabe, Y., Chaoui, A., Pingault, V., Borrego, S., and Bondurand, N. (2013). An impairment of long distance SOX10 regulatory elements underlies isolated Hirschsprung disease. Hum. Mutat. *35*, 303–307.

Lee, E. C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D. A., Court, D. L., Jenkins, N.A., and Copeland, N. G. (2001). A Highly Efficient Escherichia coli-Based Chromosome Engineering System Adapted for Recombinogenic Targeting and Subcloning of BAC DNA. Genomics *73*, 56–65.

Lee, H. J., Kweon, J., Kim, E., Kim, S., and Kim, J. (2012). Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. Genome Res. *22*, 539–548.

Lee, J. H., Skowron, P. M., Rutkowska, S. M., Hong, S. S., and Kim, S. C. (1996). Sequential amplification of cloned DNA as tandem multimers using class-IIS restriction enzymes. Genet. Anal. Biomol. Eng. *13*, 139–145.

Lee, S., Wang, W., and Liu, P. (2013). Construction of Gene Targeting Vectors by Recombineering. Methods Mol. Biol. 1–11.

Lee, S., Lee, J., Noh, K., Choi, W., Jeon, S., Taeg, G., and Kim-ha, J. (2017). Intragenic CpG islands play important roles in bivalent chromatin assembly of developmental genes.

Leeb, M., Walker, R., Mansfield, B., Nichols, J., Smith, A., and Wutz, A. (2012). Germline potential of parthenogenetic haploid mouse embryonic stem cells. Development *139*, 3301–3305.

Lenstra, T. L., Benschop, J. J., Kim, T., Schulze, J. M., Brabers, N. A. C. H., Margaritis, T., van de Pasch, L. A. L., van Heesch, S. A. A. C., Brok, M. O., Groot Koerkamp, M. J. A., et al. (2011). The Specificity and Topology of Chromatin Interaction Pathways in Yeast. Mol. Cell *42*, 536–549.

Lettice, L. A., Hill, A. E., Devenney, P. S., and Hill, R. E. (2008). Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. Hum. Mol. Genet. *17*, 978–985.

Lettice, L. A., Heaney, S., Purdie, L., Li, L., de Beer, P., Oostra, B., Goode, D., Elgar, G., Hill, R., and de Graff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Ley, T. J., Ding, L., Walter, M. J., Mclellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J., et al. (2010). Mutations in Acute

Myeloid Leukemia. N. Engl. J. Med. *363*, 2424–2433.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E. and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes & development, *24*(4), 368-380.

Li, K., Wang, G., Andersen, T., Zhou, P., and Pu, W. T. (2014). Optimization of genome engineering approaches with the CRISPR/Cas9 system. PLoS One *9*.

Li, T., Huang, S., Jiang, W. Z., Wright, D., Spalding, M. H., Weeks, D. P., and Yang, B. (2011). TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res. *39*, 359–372.

Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., Lv, J., Xie, X., Chen, Y., Li, Y., et al. (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein Cell *6*, 363–372.

Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A. M., Akopian, V., Gifford, C. A., Donaghey, J., Galonska, C., et al. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat. Genet. *47*, 469–478.

Liebhaber, S. A., Cash, F. E., and Ballas, S. K. (1986). Human α-Globin Gene Expression: The dominant role of the α-2 locus in mRNA and protein synthesis. J. Biol. Chem. *261*, 15327–15333.

Lim, W.F., Inoue-Yokoo, T., Tan, K. S., Lai, M., and Sugiyama, D. (2013). Hematopoietic cell differentiation from embryonic and induced pluripotent stem cells. Stem Cell Res. Ther. *4*, 71.

Liu, C. (2013). Strategies for Designing Transgenic DNA Constructs. Methods Mol. Biol. *1027*, 1–16.

Longo, L., Bygrave, A., Grosveld, F. G., and Pandolfi, P. P. (1997). The chromosome make-up of mouse embryonic stem cells is predictive of somatic and germ cell chimaerism. Transgenic Res. *6*, 321–328.

Lorzadeh, A., Bilenky, M., Hammond, C., Knapp, D. J. H. F., Li, L., Miller, P. H., Carles, A., Heravi-Moussavi, A., Gakkhar, S., Moksa, M., et al. (2016). Nucleosome Density ChIP-Seq Identifies Distinct Chromatin Modification Signatures Associated with MNase Accessibility. Cell Rep. *17*, 2112–2124.

Lynch, M. D., Smith, A. J. H., De Gobbi, M., Flenley, M., Hughes, J. R., Vernimmen, D., Ayyub, H., Sharpe, J. A., Sloane-Stanley, J. A., Sutherland, L., et al. (2011). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J. *31*, 317–329.

Ma, Y. D., Lugus, J. J., Park, C., and Choi, K. (2008). Differentiation of Mouse Embryonic Stem Cells into Blood. Curr. Protoc. Stem Cell Biol. 1–19.

Maeder, M. L., Angstman, J. F., Richardson, M. E., Linder, S. J., Cascio, V. M., Tsai, S. Q., Ho, Q. H., Sander, J. D., Reyon, D., Bernstein, B. E., et al. (2013a). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. *31*, 1137–1142.

Maeder, M. L., Linder, S. J., Reyon, D., Angstman, J. F., Fu, Y., Sander, J. D., and Joung, J.K. (2013b). Robust, synergistic regulation of human gene expression using TALE activators. Nat. Methods *10*, 243–245.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L., and Church, G. M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. *31*, 833–838.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., Dicarlo, J. E., Norville, J. E., and Church, G. M. (2013b). RNA-Guided Human Genome Engineering via Cas9. Science (80-.). *339*, 823–827.

Mandal, P. K., Ferreira, L. M. R., Collins, R., Meissner, T. B., Boutwell, C. L., Friesen, M., Vrbanac, V., Garrison, B. S., Stortchevoi, A., Bryder, D., et al. (2014). Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. Cell Stem Cell *15*, 643–652.

Mani, R. S., and Chinnaiyan, A. M. (2010). Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. Nat. Rev. Genet. *11*, 819–829.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., and Quon, D. (1978). The isolation of structural genes from libraries of eukaryotic DNA. Cell *15*, 687–701.

Mantsoki, A., Devailly, G., and Joshi, A. (2015). CpG island erosion, polycomb occupancy and sequence motif enrichment at bivalent promoters in mammalian embryonic stem cells. Sci. Rep. *5*.

Mantsoki, A., Devailly, G., and Joshi, A. (2016). Gene expression variability in mammalian embryonic stem cells using single cell RNA-seq data. Comput. Biol. Chem. *63*, 52–61.

Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury, W. J., Voigt, P., Martin, S. R., Taylor, W. R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U. S. A. *78*, 7634–7638.

Martins, S. B., Rino, J., Carvalho, T., Carvalho, C., Yoshida, M., Klose, J. M., De Almeida, S. F., and Carmo-Fonseca, M. (2011). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3 ′ end of human genes. Nat. Struct. Mol. Biol.

18, 1115–1123.

Maurisse, R., De Semir, D., Emamekhoo, H., Bedayat, B., Abdolmohammadi, A., Parsi, H., and Gruenert, D.C. (2010). Comparative transfection of DNA into primary and transformed mammalian cells from different lineages. BMC Biotechnol. *10*, 9.

McGrath, K., and Palis, J. (2008). Ontogeny of erythropoiesis in the mammalian embryo. Curr. Top. Dev. Biol. *82*, 1–22.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. *41*, 597–600.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., Ku, M., and Bernstein, B. E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. *6*, e1001244.

Mendenhall, E. M., Williamson, K. E., Reyon, D., Zou, J. Y., Ram, O., Joung, J.K., and Bernstein, B. E. (2013). Locus-specific editing of histone modifications at endogenous enhancers. Nat. Biotechnol. *31*, 1133–1136.

Migliaccio, A. R., Whitsett, C., Papayannopoulou, T., and Sadelain, M. (2012). The potential of stem cells as an *in vitro* source of red blood cells for transfusion. Cell Stem Cell *10*, 115–119.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J., et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. *29*, 143–148.

Miller, S. A., Mohn, S. E., and Weinmann, A. S. (2010). Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. Mol. Cell *40*, 594–605.

Minoux, M., Holwerda, S., Vitobello, A., Kitazawa, T., Kohler, H., Stadler, M. B., and Rijli, F. M. (2017). Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. Science (80-.). *355*.

Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M., and Schübeler, D. (2008). Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. Mol. Cell *30*, 755–766.

Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. Trends Biochem. Sci. *35*, 323–332.

Morin, R. D., Johnson, N. a, Severson, T. M., Mungall, A. J., An, J., Goya, R., Paul, J. E., Boyle, M., Woolcock, B. W., Kuchenbauer, F., et al. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat. Genet. *42*, 181–185.

Morita, S., Kojima, T., and Kitamura, T. (2000). Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene Ther. *7*, 1063–1066.

Moscou, M., and Bogdanove, A. (2009). A Simple Cipher Governs DNA Recognition by TAL Effectors. Science *326*, 1501.

Muramoto, T., Müller, I., Thomas, G., Melvin, A., and Chubb, J. R. (2010). Methylation of H3K4 Is Required for Inheritance of Active Transcriptional States. Curr. Biol. *20*, 397–406.

Muyrers, J. P., Zhang, Y., and Stewart, A. F. (2001). Techniques: Recombinogenic engineering–new options for cloning and manipulating DNA. Trends Biochem. Sci. *26*, 325–331.

Muyrers, J. P. P., Zhang, Y., Buchholz, F., and Stewart, A. F. (2000). RecE / RecT and Red a/ Red b specifically interacting with their respective partners initiate double-stranded break repair by specifically interacting with their respective partners. Genes Dev. *14*, 1971–1982.

Nicholls, R. D., Fischel-Ghodsian, N., and Higgs, D. R. (1987). Recombination at the human alpha-globin gene cluster: sequence features and topological constraints. Cell *49*, 369–378.

Nimmo, R. A., May, G. E., and Enver, T. (2015). Primed and ready: Understanding lineage commitment through single cell analysis. Trends Cell Biol. *25*, 459–467.

Nord, A. S., Blow, M. J., Attanasio, C., Akiyama, J. A., Holt, A., Hosseini, R., Phouanenavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., et al. (2013). Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. Cell *155*, 1521–1531.

Ntziachristos, P., Tsirigos, A., Welstead, G., Trimarchi, T., Bakogianni, S., Xu, L., Loizou, E., Holmfeldt, L., Strikoudis, A., King, B., et al. (2014). Contrasting roles for histone 3 lysine 27 demethylases in acute lymphoblastic leukemia. Nature *514*, 513–517.

O' Carroll, D., Erhardt, S., Pagani, M., Barton, S. C., Surani, M. A., Jenuwein, T. (2001). The Polycomb-Group Gene Ezh2 Is Required for Early Mouse Development. Mol. Cell. Biol. *21*, 4330–4336.

O'Geen, H., Ren, C., Nicolet, C. M., Perez, A. A., Halmai, J., Le, V. M., Mackay, J. P., Farnham, P. J., and Segal, D. J. (2017). dCas9-based epigenome editing suggests

acquisition of histone methylation is not sufficient for target gene repression. Nucleic Acids Res. *44*, 4123–4133.

Olivier, E. N., Qiu, C., Velho, M., Hirsch, R. E., and Bouhassira, E. E. (2006). Large-scale production of embryonic red blood cells from human embryonic stem cells. Exp. Hematol. *34*, 1635–1642.

Origa, R., Sollaino, M. C., Borgna-Pignatti, C., Piga, A., Feliu Torres, A., Masile, V., and Galanello, R. (2014). Alpha-globin gene quadruplication and heterozygous beta-thalassemia: A not so rare cause of thalassemia intermedia. Acta Haematol. *131*, 162–164.

Osoegawa, K., Mammoser, A. G., Wu, C., Frengen, E., Zeng, C., Catanese, J. J., and de Jong, P. J. (2001). A bacterial artificial chromosome library for sequencing the complete human genome. Genome Res. *11*, 483–496.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. Cell *152*, 157–171.

Padgett, K. A., and Sorge, J. A. (1996). Creating seamless junctions independent of restriction sites in PCR cloning. Gene *168*, 31–35.

Panzenbock, B., Bartunek, P., Mapara, M. Y., and Zenke, M. (1998). Growth and differentiation of human stem cell factor/erythropoietin-dependent erythroid progenitor cells in vitro. Blood *92*, 3658–3668.

Pasini, D., Bracken, A. P., Jensen, M. R., Denchi, E. L., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. EMBO J. *23*, 4061–4071.

Pasini, D., Cloos, P. A. C., Walfridsson, J., Olsson, L., Bukowski, J. P., Johansen, J. V, Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306–310.

Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., and Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. *31*, 839–843.

Peabody, D. S. (1993). The RNA binding site of bacteriophage MS2 coat protein. EMBO J. *12*, 595–600.

Peng, J. C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell *139*, 1290–1302.

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature *444*, 499–

502.

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., and Bejerano, G. (2013). Enhancers: five essential questions. Nat. Rev. Genet. *14*, 288–295.

Perez-Pinera, P., Ousterout, D. G., Brunger, J. M., Farin, A. M., Glass, K. A., Guilak, F., Crawford, G. E., Hartemink, A. J., and Gersbach, C. A. (2013). Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat. Methods *10*, 239–242.

Philipsen, S., and Hardison, R. C. (2017). Evolution of hemoglobin loci and their regulatory elements. Blood Cells, Mol. Dis.

Pluthero, F. G. (1993). Rapid purification of high-activity Taq DNA polymerase. Nucleic Acids Res. *21*, 4850–4851.

Preger-Ben Noon, E., Davis, F. P., and Stern, D. L. (2016). Evolved Repression Overcomes Enhancer Robustness. Dev. Cell *39*, 572–584.

Proudfoot, N. J., and Maniatis, T. (1980). The structure of a human α-globin pseudogene and its relationship to α-globin gene duplication. Cell *21*, 537–544.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Qin, J. Y., Zhang, L., Clift, K. L., Hulur, I., Xiang, A. P., Ren, B. Z., and Lahn, B. T. (2010). Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. PLoS One *5*, 3–6.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013a). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013b). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380–1389.

Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P., and Hendrich, B. (2012). NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. EMBO J. *31*, 593–605.

Rhoads, A., and Au, K. F. (2015). PacBio Sequencing and Its Applications. Genomics, Proteomics Bioinforma. *13*, 278–289.

Rickels, R., Hu, D., Collings, C. K., Woodfin, A. R., Piunti, A., Mohan, M., Herz, H. M., Kvon, E., and Shilatifard, A. (2016). An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements Implemented by Trx/MLL/COMPASS. Mol. Cell *63*, 318–328.

Rivenbark, A. G., Stolzenburg, S., Beltran, A. S., Yuan, X., Rots, M. G., Strahl, B. D., and Blancafort, P. (2012). Epigenetic reprogramming of cancer cells via targeted DNA methylation. Epigenetics *7*, 350–360.

da Rocha, S. T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N. R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. Mol. Cell *53*, 301–316.

Rossetto, D., Avvakumov, N., and Cote, J. (2012). Histone phosphorylation: A chromatin modification involved in diverse nuclear events. Epigenetics *7*, 1098–1108.

Sabo, P. J., Hawrylycz, M., Wallace, J. C., Humbert, R., Yu, M., Shafer, A., Kawamoto, J., Hall, R., Mack, J., Dorschner, M. O., et al. (2004). Discovery of functional noncoding elements by digital analysis of chromatin structure. Proc. Natl. Acad. Sci. U. S. A.*101*, 16837–16842.

Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., et al. (2006). Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nat. Methods *3*, 511–518.

Sadeh, R., Launer-Wachs, R., Wandel, H., Rahat, A., and Friedman, N. (2016). Elucidating Combinatorial Chromatin States at Single-Nucleosome Resolution. Mol. Cell *63*, 1079–1088.

Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. *32*, 347–355.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407–411.

Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juárez, P., Fernández-del-Carmen, A., Granell, A., and Orzaez, D. (2011). GoldenBraid: An iterative cloning system for standardized assembly of reusable genetic modules. PLoS One *6*.

Sauer, B., and McDermott, J. (2004). DNA recombination with a heterospecific Cre homolog identified from comparison of the pac-c1 regions of P1-related phages. Nucleic Acids Res. *32*, 6086–6095.

Savova, V., Chun, S., Sohail, M., McCole, R. B., Witwicki, R., Gai, L., Lenz, T. L., Wu, C., Sunyaev, S. R., and Gimelbrant, A. A. (2016). Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. Nat. Genet. *48*, 231–237.

Sawyer, S. T., Krantzs, S. B., and Goldwasserll, E. (1987). Binding and Receptor-mediated Endocytosis of Erythropoietin in Friend Virus-Infected Erythroid Cells. J. Biol. Chem. *262*, 5554–5562.

Schlake, T., and Bode, J. (1994). Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. Biochemistry *33*, 12746–12751.

Schmitges, F. W., Prusty, A. B., Faty, M., Stützer, A., Lingaraju, G. M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell *42*, 330–341.

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. Nat. Protoc. *3*, 1101–1108.

Schmitz, M. L., and Baeuerle, P. A. (1991). The p65 subunit is responsible for the strong transcription activating potential of NF-kappa B. EMBO J. *10*, 3805–3817.

Seferos, D. S., Giljohann, D. A., Hill, H. D., Prigodich, A. E., and Mirkin, C. A. (2007). Nano-flares: Probes for transfection and mRNA detection in living cells. J. Am. Chem. Soc. *129*, 15477–15479.

Sen, S., Block, K. F., Pasini, A., Baylin, S. B., and Easwaran, H. (2016). Genome-wide positioning of bivalent mononucleosomes. BMC Med. Genomics *9*, 1–14.

Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., and Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat. Biotech. *22*, 1567–1572.

Shema, E., Jones, D., Shoresh, N., Donohue, L., Ram, O., and Bernstein, B. E. (2016). Single-molecule decoding of combinatorially modified nucleosomes. Science. *352*, 717–721.

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu. Rev. Biochem. *81*, 65–95.

Simon, J. A., and Kingston, R. E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat. Rev. Mol. Cell Biol. *10*, 697–708.

Slack, J. M. W. (2002). Conrad Hal Waddington: the last Renaissance biologist? Nat. Rev. Genet. *3*, 889–895.

Sollaino, M. C., Paglietti, M. E., Loi, D., Congiu, R., Podda, R., and Galanello, R. (2010). Homozygous deletion of the major alpha-globin regulatory element (MCS-R2) responsible for a severe case of hemoglobin H disease. Blood *116*, 2193–2194.

Soltis, R. D., Hasz, D., Morris, M. J., and Wilson, I. D. (1979). The effect of heat inactivation of serum on aggregation of immunoglobulins. Immunology *36*, 37–45.

Son, J., Shen, S. S., Margueron, R., and Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes & Development*, *27*(24), 2663–2677.

Souroullas, G. P., Jeck, W. R., Parker, J. S., Simon, J. M., Liu, J., Paulk, J., Xiong, J., Clark, K. S., Fedoriw, Y., Qi, J., et al. (2016). An oncogenic Ezh2 mutation induces tumors through global redistribution of histone 3 lysine 27 trimethylation. Nat. Med. *22*, 1–15.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–517.

Spitz, F., and Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Stepper, P., Kungulovski, G., Jurowska, Reneta, Chandra, T., Krueger, F., Reinhardt, R., Reik, W., Jeltsch, A., and Jurkowski, T. (2016). Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. Nucleic Acids Res. 1–15.

Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. *9*, 1428–1435.

Streubel, J., Baum, H., Grau, J., Stuttman, J., and Boch, J. (2017). Dissection of TALE-dependent gene activation reveals that they induce transcription cooperatively and in both orientations. PLoS One *12*, 1–24.

Subach, O. M., Gundorov, I. S., Yoshimura, M., Subach, F. V., Zhang, J., Grüenwald, D., Souslova, E. A., Chudakov, D. M., and Verkhusha, V. V. (2008). Conversion of Red Fluorescent Protein into a Bright Blue Probe. Chem. Biol. *15*, 1116–1124.

Sugawara, A., Goto, K., Sotomaru, Y., Sofuni, T., and Ito, T. (2006). Current Status of Chromosomal Abnormalities in Mouse Embryonic Stem Cell Lines Used in Japan. Comp. Med. *56*, 31–34.

Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. Science *332*, 472–474.

Suzuki, H., Kamada, N., Ueda, O., Jishage, K., Kurihara, Y., Kurihara, H., Terauchi, Y., Azuma, S., Kadowaki, T., Kodama, T., et al. (1997). Germ-Line Contribution of Embryonic Stem Cells in Chimeric Mice: Influence of Karyotype and In Vitro Differentiation Ability. Exp. Anim. *46*, 17–23.

Symington, L. S., and Gautier, J. (2011). Double-Strand Break End Resection and Repair Pathway Choice. Annu. Rev. Genet. *45*, 247–271.

Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E.F., and Vignali, D. A. A. (2004). Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat. Biotechnol. *22*, 589–594.

Tai, D. J. C., Ragavendran, A., Manavalan, P., Stortchevoi, A., Seabra, C.M., Erdin, S., Collins, R. L., Blumenthal, I., Chen, X., Shen, Y., et al. (2016). Engineering microdeletions and microduplications by targeting segmental duplications with CRISPR. Nat. Neurosci. *19*, 517–522.

Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F., and Greaves, D. R. (1989). A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. Nature *338*, 352–355.

Tamm, C., Galitó, S.P., and Annerén, C. (2013). A comparative study of protocols for mouse embryonic stem cell culturing. PLoS One *8*, 1–10.

Tamm, C., Kadekar, S., Pijuan-Galitó, S., and Annerén, C. (2016). Fast and efficient transfection of mouse embryonic stem cells using non-viral reagents. Stem Cell Rev. *12*, 584–591.

Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S., and Vale, R. D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Gayther, S. A., Apostolidou, S., Jones, A., Lechner, M., Beck, S., Jacobs, I. J., et al. (2009). An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS One *4*, e8274.

Thakore, P. I., Black, J. B., Hilton, I. B., and Gersbach, C. A. (2016). Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nat. Methods *13*, 127–137.

Therizols, P., Illingworth, R. S., Courilleau, C., Boyle, S., Wood, A. J., and Bickmore, W. A. (2014). Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. Science *346*, 1238–1242.

Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo- derived stem cells. Cell *51*, 503–512.

Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R., Deaton, A., Andrews, R., James, K. D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082–1086.

Torres-Padilla, M. E. E., and Chambers, I. (2014). Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage [review]. Development *141*, 2173–2181.

Trakarnsanga, K., Griffiths, R. E., Wilson, M. C., Blair, A., Satchwell, T. J., Meinders, M., Cogan, N., Kupzig, S., Kurita, R., Nakamura, Y., et al. (2017). An immortalized adult human erythroid line facilitates sustainable and scalable generation of functional red cells. Nat. Commun. *8*.

Trojer, P., and Reinberg, D. (2007). Facultative Heterochromatin: Is There a

Distinctive Molecular Signature? Mol. Cell *28*, 1–13.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. (2012). Primer3-new capabilities and interfaces. Nucleic Acids Res. *40*, 1–12.

Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., and Holmes, M. C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646–651.

Vastenhouw, N. L., Zhang, Y., Woods, I. G., Imam, F., Regev, A., Liu, X. S., Rinn, J., and Schier, A. F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature *464*, 922–926.

Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W. M. P., van Schaik, F. M. A., Varier, R. A., Baltissen, M. P. A., Stunnenberg, H. G., Mann, M., and Timmers, H. T. M. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58–69.

Vernimmen, D. (2014). Uncovering Enhancer Functions Using the α-Globin Locus. PLoS Genet. *10*.

Vernimmen, D., and Bickmore, W. A. (2015). The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. Trends Genet. *31*, 696–708.

Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G., and Higgs, D. R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J. *26*, 2041–2051.

Vernimmen, D., Marques-Kranc, F., Sharpe, J. A., Sloane-Stanley, J. A., Wood, W. G., Wallace, H. A. C., Smith, A. J. H., and Higgs, D. R. (2009). Chromosome looping at the human alpha-globin locus is mediated via the major upstream regulatory element (HS -40). Blood *114*, 4253–4260.

Vernimmen, D., Lynch, M.D., De Gobbi, M., Garrick, D., Sharpe, J. a, Sloane-Stanley, J. a, Smith, A.J.H., and Higgs, D.R. (2011). Polycomb eviction as a new distant enhancer function. Genes Dev. *25*, 1583–1588.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J. T., Jasinska, A. J., et al. (2015). Enhancer Evolution across 20 Mammalian Species. Cell *160*, 554–566.

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J. M., et al. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871–874.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

Voigt, P., LeRoy, G., Drury, W. J., Zee, B. M., Son, J., Beck, D. B., Young, N. L., Garcia, B. A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181–193.

Voigt, P., Tee, W. W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes Dev. *27*, 1318–1338.

Voit, R. A., Hendel, A., Pruett-Miller, S. M., and Porteus, M. H. (2014). Nuclease-mediated gene editing by homologous recombination of the human globin locus. Nucleic Acids Res. *42*, 1365–1378.

Wallace, H. A. C., Marques-Kranc, F., Richardson, M., Luna-Crespo, F., Sharpe, J. A., Hughes, J., Wood, W. G., Higgs, D. R., and Smith, A. J. H. (2007). Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. Cell *128*, 197–209.

Walport, L. J., Hopkinson, R. J., Vollmar, M., Madden, S. K., Gileadi, C., Oppermann, U., Schofield, C. J., and Johansson, C. (2014). Human UTY(KDM6C) is a male-specific Nϵ-methyl lysyl demethylase. J. Biol. Chem. *289*, 18302–18313.

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

Wang, L., Brown, J. L., Cao, R., Zhang, Y., Kassis, J. A., Jones, R. S., Hill, C., and Carolina, N. (2004). Hierarchical Recruitment of Polycomb Group Silencing Complexes. Mol. Cell *14*, 637–646.

Wang, R., Clark, R., and Bautch, V. L. (1992). Embryonic stem cell-derived cystic embryoid bodies form vascular channels: an in vitro model of blood vessel development. Development *114*, 303–316.

Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Peng, W., Zhang, M. Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897–903.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. PLoS One *6*.

Weiner, A., Lara-Astiaso, D., Krupalnik, V., Gafni, O., David, E., Winter, D.R., Hanna, J. H., and Amit, I. (2016). Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution. Nat. Biotechnol. *34*, 953–961.

Weiss, I. M., and Liebhaber, S. A. (1995). Erythroid cell-specific mRNA stability elements in the alpha 2-globin 3' nontranslated region. Mol. Cell. Biol. *15*, 2457–2465.

Weiss, I. M., and Liebhaber, S. A. (1994). Erythroid cell-specific determinants of

alpha-globin mRNA stability. Mol. Cell. Biol. *14*, 8123–8132.

Weiss, M. J., and Orkin, S. H. (1996). *In Vitro* Differentiation of Murine Embryonic Stem Cells New Approaches to Old Problems. J. Clin. Invest. *97*, 1999–2002.

Wozniak, G. G., and Strahl, B. D. (2014). Hitting the 'mark': interpreting lysine methylation in the context of active transcription. Biochim. Biophys. Acta *1839*, 1353–1361.

Wu, M. Y., He, Y., Yan, J. M., and Li, D. Z. (2016). A novel selective deletion of the major alpha-globin regulatory element (MCS-R2) causing alpha-thalassaemia. Br. J. Haematol. *176*, 984–986.

Wu, X., Johansen, J. V., and Helin, K. (2013). Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. Mol. Cell *49*, 1134–1146.

Wu, X., Kriz, A. J., and Sharp, P. A. (2014). Target specificity of the CRISPR-Cas9 system. Quant. Biol. *2*, 59–70.

Wyman, C., and Kanaar, R. (2006). DNA double-strand break repair: all's well that ends well. Annu. Rev. Genet. *40*, 363–383.

Xiao, T., Hall, H., Kizer, K. O., Shibata, Y., Hall, M. C., Borchers, C. H., and Strahl, B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev. *17*, 654–663.

Xie, F., Ye, L., Chang, J. C., Beyer, A. I., Wang, J., Muench, M. O., and Kan, Y. W. (2014). Seamless gene correction of beta-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. Genome Res. *24*, 1526–1533.

Yang, H., Wang, H., Shivalila, C. S., Cheng, A. W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370–1379.

Yang, L., Rau, R., and Goodell, M. A. (2015). DNMT3A in haematological malignancies. Nat. Rev. Cancer *15*, 152–165.

Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Yoshimi, K., Kunihiro, Y., Kaneko, T., Nagahora, H., Voigt, B., and Mashimo, T. (2016). ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. Nat. Commun. *7*, 10431.

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *97*, 5978–5983.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012). Dense Chromatin Activates Polycomb Repressive Complex 2 to Regulate H3 Lysine 27 Methylation. Sci. Transl. Med. *337*, 971–975.

Yusa, K., Rashid, S. T., Strick-Marchand, H., Varela, I., Liu, P. Q., Paschon, D. E., Miranda, E., Ordóñez, A., Hannan, N. R. F., Rouhani, F. J., et al. (2011). Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. Nature *478*, 391–394.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., et al. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. *38*, 4246–4253.

Zhang, Y., Vastenhouw, N. L., Feng, J., Fu, K., Wang, C., Ge, Y., Pauli, A., Van Hummelen, P., Schier, A. F., and Liu, X. S. (2014). Canonical nucleosome organization at promoters forms during genome activation. Genome Res. *24*, 260–266.

Zhen, C. Y., Tatavosian, R., Huynh, T. N., Duc, H. N., Das, R., Kokotovic, M., Grimm, J. B., Lavis, L. D., Lee, J., Mejia, F. J., et al. (2016). Live-cell single-molecule tracking reveals co-recognition of H3K27me3 and DNA targets polycomb Cbx7-PRC1 to chromatin. Elife *5*, e17667.

Zheng, L., Baumann, U., and Reymond, J. L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res. *32*, e115.

Zheng, X., Yue, S., Chen, H., Weber, B., Jia, J., and Zheng, Y. (2015). Low-Cell-Number Epigenome Profiling Aids the Study of Lens Aging and Hematopoiesis. Cell Rep. *13*, 1505–1518.

Zhong, C., Yin, Q., Xie, Z., Bai, M., Dong, R., Tang, W., Xing, Y.H., Zhang, H., Yang, S., Chen, L. L., et al. (2015). CRISPR-Cas9-Mediated Genetic Screening in Mice with Haploid Embryonic Stem Cells Carrying a Guide RNA Library. Cell Stem Cell *17*, 221–232.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation. Mol. Cell *29*, 69–80.

Chapter 9 Appendices

9.1 Primer sequences used in cloning experiments

Table 0.1 List of cloning primer sequences used in Chapter 4.

Figure Reference	Primer Name	Sequence 5' - 3'
4.3A	Halpha2_Gent_Fw	TGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAATACCGTGCGTGTTTCGAGCATGTTTCTGCGTAGTGTCAGCTCATCC
4.4A, B	HumAlpha2UTR_Gent_Rev_NEW	CGTTGGGAGGCCCAGCGGGCAGGAGGAACGGCTACCGAGGCTCCAGCTTAGTGTCCATCCTGTAGGTGTAGACGACGACGAACAGAG
4.3B	Halpha2_HA_ex3_Fwd	TGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAATACCGT
	Halpha2_HA_3??UTR_Rev	CGTTGGGAGGCCCAGCGGGCAGGAGGAACGGCTACCGAGGCTCCAGCTTA
4.5A, B	a2_exon3_Fwd_Screen	ACCTCCCCGCCGAGTTC
	a2_3'UTR_Rev_Screen	ATACTCCCTGCAGTTCTCCC
4.7B	G_pAB001_Fwd	GGCGCGCCATTGATCTAG
3'arm Gibson	G_pAB001_Rev	CTTGTACAGCTCGTCCATGC
4.7B	G_3'_arm_Fwd	GGACGAGCTGTACAAGTAAGCTGGAGCCTCGGTAG
3'arm Gibson	G_3'_arm_Rev	GATCAATGGCGCGCCCAGGAGGTCCCAGAGGCC
4.7C	G_p2A-mCh_ha2_Fwd	GTGAAACAGACTTTGAATTTTGACCTTCTCTCAAGTTG
5'arm Gibson	G_p2A-mCh_ha2_Rev	GGCGCGCCATTGGATCGG
4.7C	G_5'arm_Fwd	ATCCAATGGCGCGCCGCCAGCCAATGAGCGC
5'arm Gibson	G_5'arm_Rev	ATTCAAAGTCTGTTTCACACGGTATTTGGAGGTCAGCACG
4.8A	Gib_pAB-002_F_13	CAGGAGGTCCCAGAGGCC
3'arm ext Gibson	Gib_pAB-002_R_13	GGCGCGCCATTGATCTAG
4.8A	Gib_3armext_Fwd	CTCTGGGACCTCCTGGTCTTCTGCTTCCTGTGC
3'arm ext Gibson	Gib_3armext_Rev	GATCAATGGCGCGCCTGCCCGGCTGATTTTGTA
4.9A	Gib_pAB-013_Rev	CTGGGTTCAACCAGGGAG
PB:Puro Gibson	Gib_pAB-013_Fwd	CATACGCTCTCCATCAAAAC
4.9A	Gib_PB_puroTK_F	CCTGGTTGAACCCAGTTAACCCTAGAAAGATAGTCTG
PB:Puro Gibson	Gib_PB_puroTK_R	TGATGGAGAGCGTATGTTAACCCTAGAAAGATAATCATATTG
4.10B	A.hnuc2.TOP	CACCGTAGCCGTTCCTCCTGCCCGC
sgRNA-A oligos	A.hnuc2.BOTTOM	AAACGCGGGCAGGAGGAACGGCTAC
4.10C	HBA2 exon3 TAA Fwd DV	CTGACCTCCAAATACCGTTAAGCT
T7 primers	HBA2 3UTR DV Rev4	TCCCATACTCCCTGCAGTTCTC
4.11	pAB 29 SDM Fwd2	GTTCCTcctgcccgctgggcc
WT SDM primers	pAB 29 SDM Rev2	ggctaccgaggctcca

Table 9.2 List of genotyping primer sequences used in Chapter 5.

Reference	Primer Name	Sequence 5' - 3'
PCR1: mCherry	a2_exon3_Fwd_Screen	ACCTCCCCGCCGAGTTC
	a2_3'UTR_Rev_Screen	ATACTCCCTGCAGTTCTCCC
PCR2: 5`arm	HBA2_5_Insert_2_F	TTTTTGCGTCCTGGTGTTTGT
	mCherry Rev 2	cttgctcaccatGCCTGATC
PCR3: 3`arm	PB GBlock 3arm Fwd	CTAGACCCAGCTTTCTTGTACAA
	Gen Rev Screen 2	CTCTTCTGACTCTGCCCACA
PCR4: mCherry-PB:Puro	mCherry Fwd 2	tcccacaacgaggactacac
	puroTK Rev	ACTCGGTCCCCATGGTTTAG
PCR5: PB:Puro	a2 3`UTR Fwd Screen	GGCCCTTCCTGGTCTTTGAA
	HBA2_5_Insert_Rev3	GAGGAGGTGAGACTTAAGGATATGT
5` Southern blot Probe PCR	5' Probe Fwd	TTTGTAAGGTGCATTCAGAACTCA
	5' Probe Rev	CCAAAGCTCAATTCAGTACTGTGA

Table 0.2 List of cloning primers used in Chapter 6.

Figure Reference	Primer Name	Sequence 5' - 3'
6.6Ai	RbBP5_Fwd	ATGAACCTCGAGTTGCTGGAG
Rbbp5 cloning	RbBP5_Rev	TCACAGCAGTTCTGAGATGGC
6.6Aii	M_JMJD3_JmjC_Fwd	CGAGAGTCCTACCTGTCC
Jmjd3 cloning	JMJD3_JmjC_Rev_New	TCATCGAGACGTGCTGGC
6.6Aiii	EZH2_cDNA_Fwd	ATGGGCCAGACTGGGAAGAAATC
Ezh2 cloning	EZH2_cDNA_Rev	TCAAGGGATTTCCATTTCTCG
6.7B	267 3x FLAG TOP	ctagaGACTACAAGGACCACGACGGTGACTACAAGGACCACGAC- ATCGACTACAAGGACGACGACGACAAGTGATAAGTGAGACGATAgc
3xFLAG cloning	268 3x FLAG BOTTOM	ggccgcTATCGTCTCACTTATCACTTGTCGTCGTCGTCCTTGTAG- TCGATGTCGTGGTCCTTGTAGTCACCGTCGTGGTCCTTGTAGTCt
TALE Only	269 3x FLAG NO GOI TOP	ctagaGTACGTCTCAGGTGGTGGTTCTGACTACAAGGACCACGACGGTGACTACAA- GGACCACGACATCGACTACAAGGACGACGACGACAAGTGATAAGTGAGACGATAgc
3xFLAG cloning	270 3x FLAG NO GOI BOTTOM	ggccgcTATCGTCTCACTTATCACTTGTCGTCGTCGTCCTTGTAGTCGATGTCGT- GGTCCTTGTAGTCACCGTCGTGGTCCTTGTAGTCAGAACCACCACCTGAGACGTACt
6.8A	262 PGK Fwd pStb205	CTGtctagaCAGGTCGTGAGACG*aattctaccgggtaggggagg*
PGK cloning	271 PGK Rev New pStb205	CTTGATACAGgcggccgcTATGTACGTCTCTggta*tgcaggtcgaaaggcccg*
6.10Ai	255 RBBP5 FWD FLAG	ctgGAATTCGTACGTCTCAGGTGGTGGTTCT*AACCTCGAGTTGCTGGAGTCC*
Rbbp5 into Kan vector	256 RBBP5 REV FLAG	ttcTCTAGA*CAGCAGTTCTGAGATGGCtcc*
6.10Aii	257 JMJD3 FWD FLAG	ctgGAATTCGTACGTCTCAGGTGGTGGTTCT*TACCTGTCCCCTGCCC*
Jmjd3 into Kan vector	258 JMJD3 REV FLAG	ttcTCTAGA*TCGAGACGTGCTGGCG*
6.10Aiii	274 EZH2 XmaI 3x Flag fwd	ctgGGGCCCGTACGTCTCAGGTGGTGGTTCT*GGCCAGACTGGGAAGA*
Ezh2 into Kan vector	275 EZH2 XmaI 3x Flag rev	ttcGGGCCCAGGGATTTCCATTTCTCG
6.10Aii	293 EZH2 SDM Fwd	tctagaGACTACAAGGACCACGAC
Fix ORF in Ezh2	294 EZH2 SDM Rev	AGGGATTTCCATTTCTCGTT
6.12Aii	262 PGK Fwd pStb205	CTGtctagaCAGGTCGTGAGACG*aattctaccgggtaggggagg*
PGK colony PCR	271 PGK Rev New pStb205	CTTGATACAGgcggccgcTATGTACGTCTCTggta*tgcaggtcgaaaggcccg*
6.13Aii	249 JMJD3 Fwd Screen	ACCACCATCGCTAAATACGC
Jmjd3 colony PCR	250 JMJD3 Rev Screen	CCACCAGGAACCAGTCAAGT
6.13Aiii	EZH2 Fwd Screen	AAGAACACAGAAACAGCTTTGGA
Ezh2 colony PCR	248 EZH2 Rev Screen	GACAGAGGTCAGGGTCACAC
6.13Aiv	TalC_Fwd	TCAGAAGAGTCAATCGCCGT
TALE only colony PCR	TalC_3'Rev	AGTGACCTGTTCGTTGCAAC

9.2 Compatibility of overhangs produced in Golden Gate cloning experiments

Table 9.4 List of *Bsm*BI sites and overhangs in Golden Gate cloning plasmids used in Chapter 6. Each *Bsm*BI restriction site (red) and the overhangs that are generated (teal) are shown for all of the Golden Gate plasmids used in Chapter 6. No two overhangs are compatible with any site other than their intended fusion partner, thus ensuring correct ligation of each module.

BsmBI site
BsmBI overhang

5'	CGTCTCN		NNNNNGAGACG
3'	GCAGAGNNNNN		NCTCTGC

SEGMENT	SIDE	OVERHANG	OVERHANG		SEQUENCE	
pTFE	5'	5'	CAGC	5'	CGTCTTGTCGACTCGTCTCTTACC	3'
Promoter	3'	5'	TACC	3'	GCAGAACAGCTGAGCAGAGAATGG	5'
PCR	5'	5'	GTCG	5'	GTCGTGAGACG...TACCAGAGACG	3'
PROMOTER	3'	5'	ATGG	3'	CAGCACTCTGC...ATGGTCTCTGC	5'
pTFE	5'	5'	GGGA	5'	CCCTGGAGACG...CGTCTCCAACG	3'
TALE Array	3'	5'	AACG	3'	GGGACCTCTGC...GCAGAGGTTGC	5'
pFusA	5'	5'	CCCT	5'	CCCT...TALE...	3'
	3'	5'	ACCG	3'	...TALE...ACCG	5'
pFusB	5'	5'	TGGC	5'	TGGC...TALE	3'
	3'	5'	GATA	3'	...TALE...GATA	5'
LR	5'	5'	CTAT	5'	CTAT...LR...	3'
	3'	5'	TTGC	3'	...LR...TTGC	5'
pTFE	5'	5'	CCAC	5'	GGTGGGAGACG...CGTCTCTTAAG	3'
GOI	3'	5'	TAAG	3'	CCACCCTCTGC...GCAGAGAATTG	5'
PCR	5'	5'	GGTG	5'	CGTCTCAGGTG...TAAGTGAGACG	3'
GOI	3'	5'	ATTC	3'	GCAGAGTCCAC...ATTCACTCTGC	5'
pTFE	5'	5'	AGTG	5'	TCACCGAGACG	3'
BFP	3'	5'	TCAC	3'	AGTGGCACTGC	5'
JMJD3	5'	5'	TAGA	5'	CGTCTCGATCTA	3'
	3'	5'	ATCT	3'	GCAGAGCTAGAT	5'
EZH2	5'	5'	TTCT	5'	AAGATGAGACG	3'
	3'	5'	AAGA	3'	TTCTACTCTGC	5'
PGK	5'	5'	GATC	5'	CGTCTCACTAG	3'
	3'	5'	CTAG	3'	GCAGAGTGATC	5'
STB205	5'	5'	GAGT	5'	CGTCTCGCTCAG	3'
	3'	5'	CTCA	3'	GCAGAGCGAGTC	5'
pGREEN 1	5'	5'	CCCC	5'	GGGGGGAGACG	3'
	3'	5'	GGGG	3'	CCCCCCTCTGC	5'
pGREEN 2	5'	5'	TGAG	5'	TGAGCGAGACG	3'
	3'	5'	ACTC	3'	ACTCGCTCTGC	5'

9.3 Infrequently used abbreviations

^{32}P	Phosphorus-32
3C	Chromatin conformation capture
AHSP	Alpha haemoglobin stabilising protein
ANOVA	One-way analysis of variance
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BFP	Blue fluorescent protein
bp	Base pair
BSA	Bovine serum albumin
Cas9n	Cas9 nickase
CBP	Chicken beta actin short promoter and CMV enhancer
Cfp1	CxxC finger protein 1
Cre	Causes recombination
dbSNP	Short genetic variations database
Dcp1A	Decapping messanger ribonucleic acid 1A
DMF	Dimethylformamide
DOT1L	Disruptor of telomeric silencing 1-like
DsRed	*Discosoma sp.* Red
DTT	Dithiothreitol
eGFP	Enhanced green fluorescent protein
EpiSC	Epiblast stem cells
Epo	Erythropoietin
EtOH	Ethanol
EWAS	Epigenome-wide association studies
FAM	6-carboxyfluorescein
FANTOM	Functional annotation of the mammalian genome
FIAU	Fialuridine
FISH	Fluorescent *in situ* hybridisation
FITC	Fluorescein isothiocyanate
FLP	Flippase
FRT	Flippase recombination target site
G9a	G9a histone methyltransferase
HA	Homology arms
HEK293T	Human embryonic kidney 293T
HeLa cells	Henrietta Lacks cells
hES cell	Human embryonic stem cell
Hox	Homeotic
HP1	Heterochromatin protein 1
HPRT	Hypoxanthine-guanine phosphoribosyltransferase
HR	Homologous recombination
HRP	Horseradish peroxidase
HSF1	Heat shock factor 1
HXKYacZ	Histone X lysine Y acetylated Z times

HXKYmeZ	Histone X lysine Y methylated Z times
iCpG	Intragenic CpG
IDT	Integrated DNA technologies
IMDM	Iscove's modified dulbecco's medium
IP	Immunoprecipitation
iPS cell	Induced pluripotent stem cell
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JARID	Jumonji and AT-rich interaction domain containing protein
KAP1	KRAB associated protein 1
KDM	Histone lysine demethylase
KRAB	Krüppel-associated box
LB	Lysogeny broth
LIF	Leukemia inhibitory factor
loxP	Locus of X-over P1
LR	Last half repeat
LSD1	Lysine-specific histone demethylase 1A
Mb	Megabase
MCR-R	Multispecies conserved region
MEM	Minimum essential medium
MLL	Mixed lineage leukaemia
MMEJ	Microhomology-mediated end joining
MoClo	Modular cloning
MOPS	3-(N-morpholino)propanesulfonic acid
NCBI	National Centre for Biotechnology Information
NEB	New England Biolabs Inc.
Neo	Neomycin
NF-kB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NPC	Neural precurser cell
Nrp1	Neuropilin 1
NuRD	Nuclease remodeling deacetylase
NURF	Nucleosome remodeling factor
Oct4	Octamer-binding transcription factor 4
PacBio	Pacific Biosciences
PAM	Protospacer adjacent motif
PB:PuroΔTK	PGK-puromycinΔthymidine-kinase counter-selection cassette flanked with piggyBac inverted repeats
PHD	Plant homeodomain
Plat-E	Platinum-E
PolII	RNA polymerase II complex
PRC	Polycomb repressive complex
PRDM9	Proline rich domain zinc finger protein 9
PRE	Polycomb repressive element
PVDF	Polyvinylidene difluoride
rcf	Relative centrifugal force

re-ChIP	Sequential chromatin immunoprecipitation
REST	RE1-silencing transcription factor
RNA-seq	Ribonucleic acid next generation sequencing
RNAi	Ribonucleic acid interference
rpm	Rotations per minute
RT	Recombination target site
SANT	Switching-defective protein 3, adaptor 2, nuclear receptor corepressor, transcription factor IIIB
SCF	Stem cell factor
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulfate
SET	Su(var)3-9, enhancer-of-zeste and trithorax
sgRNA	Short single guide ribonucleic acid
SM	Selection marker
SMRT	Single molecule- real time
SOB	Super optimal broth
SOC	SOB with catabolite repression
Spec	Spectinomycin
SSC	Saline-sodium citrate
ssDNA	Single stranded DNA
SSR	Site-specific recombinase
Suz12	Suppressor of zeste 12
SV40	Simian virus 40
T4 PNK	T4 Polynucleotide Kinase
TAE	Tris-acetate EDTA
TAF	TATA-box binding protein associated factor 1
TALEN	Transcription activator-like effector nuclease
TAMRA	Tetramethylrhodamine
TBE	Tris-borate EDTA
TBST	Tris-buffered saline supplemented with Tween 20
TE	Tris-EDTA
TET	Ten-eleven translocation methylcytosine dioxygenase
TFIID	Transcription factor II D
TSS	Transcription start site
Ubc	Ubiquitin C
UCSC	University of California, Santa Cruz
Utf1	Undifferentiated embryonic cell transcrption factor 1
UTX/Y	Ubiquitously transcribed tetratricopeptide repeat, X/Y
VP	Viral protein
Wdr5	WD repeat-containing protein 5
WT	Wild-type
Xist	X-inactive specific transcript
ZFN	Zinc-finger nuclease